

Vol. XVII Part 1

May, 1964

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
R. J. AUDLEY

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

C. BANKS	A. R. JONCKHEERE
M. S. BARTLETT	D. N. LAWLEY
G. H. BOWER	R. D. LUCE
D. E. BROADBENT	S. MESSICK
J. BROWN	E. A. PEEL
V. R. CANE	L. S. PENROSE
D. R. COX	J. FRASER ROBERTS
L. J. CRONBACH	W. STEPHENSON
C. I. HOWARTH	A. STUART
P. E. VERNON	

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20s. per part (U.S.A. \$3.50)

Subscription 30s. 6d. per volume (U.S.A. \$4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The *British Journal of Statistical Psychology* is issued by the British Psychological Society. The subscription price per volume is 30s. net for Members of the Society and 30s. 6d. (post free) for non-members. The subscription price in the U.S.A. is \$4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1: non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, about May and November.

Papers for publication and books for review should be sent to the Editor, R. J. Audley, Department of Psychology, University College London, Gower Street, W.C.1. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 50 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology*, at a subscription rate of 60s. per annum for either journal, and *The British Journal of Social and Clinical Psychology* three times a year at a subscription rate of 50s. per annum. Subscriptions for these journals should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology* three times a year: for this the subscription, 30s. per annum, should be sent to Methuen & Co. Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1.

Papers for publication in *The British Journal of Psychology* should be sent to Professor Arthur Summerfield, Department of Psychology, Birkbeck College, Malet Street, London, W.C.1, those for publication in *The British Journal of Medical Psychology* to Dr. Thomas Freeman, Lansdowne Clinic, 4 Royal Crescent, Kelvingrove, Glasgow, C.3; those for publication in *The British Journal of Social and Clinical Psychology* to Mr. Michael Argyle (Social Psychology Editor), Institute of Experimental Psychology, 1 South Parks Road, Oxford, or to Dr. Ralph Hetherington, Department of Studies in Psychological Medicine, 6 Abercromby Square, Liverpool 7; and those for publication in *The British Journal of Educational Psychology* to Mr. L. B. Birch, University of Sheffield, Institute of Education, Sheffield 10.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer: that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract: (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the 'Recommendations to Authors' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21-73), or to those given in the paper on 'Setting Mathematics' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost. Contributors should note that the *Journal* is intended for the general psychological reader, not solely for those who specialize in statistics. Unusual technicalities should be either avoided or carefully explained, e.g. by reference to previous discussions of the same or kindred problems in this *Journal*.

In future authors will receive one complete slip-proof of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.

All manuscripts should be submitted in duplicate.

ASYMPTOTIC LEARNING IN PSYCHOPHYSICAL THEORIES¹

By R. DUNCAN LUCE

University of Pennsylvania

The major types of models that have been proposed to account for the psychophysical data that are obtained when the stimulus differences are small are described briefly. Because it is clear that contingency variables, such as presentation schedules and payoffs, as well as the physical stimuli affect the response probabilities, recent models have included a trial dependent decision mechanism in addition to the (usually) static sensory one. Such models all appear to be special cases of two very general, but mathematically distinct, families of models, which are formulated in eqns. 1 and 5. Hypotheses about how the subject selects the response-bias parameters of the decision process are examined. The assumption that they are chosen so as to maximize the expected payoff leads to incorrect predictions for special cases of both families. The alternative hypothesis studied is that the bias parameters are altered from trial to trial on the basis of information feedback according to one or another stochastic learning model. Primary attention is paid to the asymptotic expected values predicted for these parameters. Several such learning processes are described, their relations to static psychophysical models outlined, and their ability to explain data discussed.

I. INTRODUCTION

Modern psychophysicists divide into two camps, the one concerned with people's responses to small stimulus perturbations and the other with responses to large perturbations. It is doubtful if anyone believes the present cleavage in method and theory to be a fundamental scientific distinction, but it is nevertheless real now. I shall be concerned here with the psychophysics of small differences. Included is the research on thresholds and, more generally, on the detection of stimuli that are difficult to detect, on the recognition of one of several possible stimuli, and on the discriminability of stimuli. The methods used are well known and in many cases classical: constant stimuli, single stimuli, pair comparisons, yes-no and forced-choice techniques, and so on. The methods for studying large perturbations are the (superficially) direct scaling techniques applicable to the whole sensory continuum, e.g., the methods of bisection, fractionation, magnitude estimation, category scaling, etc.

Perhaps the neatest distinction between the methods of the two camps of psychophysicists rests upon the fact that in studying small differences one can sensibly class responses as correct and incorrect in terms of physical properties of the stimuli, whereas this cannot be done with the methods used in studying large differences. This means that information feedback and payoffs can be, and often are, used in experiments of the first type but not in those of the second

¹ This paper was presented at the International Congress of Scientific Psychology, Washington, D.C., August 1963. Its preparation was supported in part by National Science Foundation Grant NSF G 17637 to the University of Pennsylvania and by the Social Science Research Council Summer Seminar on Mathematical Psychophysics.

type. For a more detailed discussion of these matters, see the chapters on psychophysics in Luce, Bush, and Galanter (1963).

An important conceptual revolution in the study of responses to small differences occurred during the past decade, and it is still affecting the collection, analysis, and theoretical interpretation of data. At the risk of repeating the familiar, let me be explicit about this development since the material covered in this paper is one of its current phases. Perhaps the revolution is best symbolized by the so-called receiver operating characteristic (ROC) curve or, to use the term I prefer, the iso-sensitivity curve. In the simplest case of a Yes-No experiment, consider the plot of the conditional probability $p(Y|s)$ of a 'yes' response, Y , to a stimulus presentation, s , against the conditional probability $p(Y|n)$ of a 'yes' response on those occasions when no stimulus is presented, n .

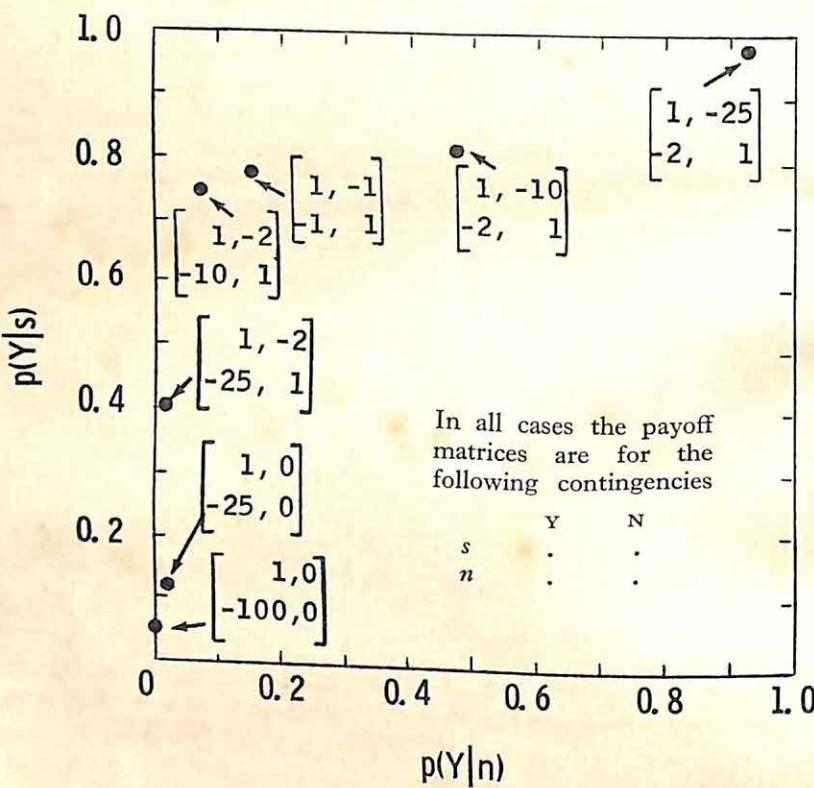


FIGURE 1. Iso-sensitivity data (subject 6) reported by Norman (1962) for fixed energy increments in a 1000 c.p.s. pure tone. The increment was 2.2% of the background energy; the presentation probability of the signal was 0.5; and the payoff matrices (in cents) are shown. Each point is estimated from 200 to 300 observations.

An iso-sensitivity curve is the locus of points that is generated when certain variables are varied but *when the stimulus conditions are held fixed*. The important observation is that such loci appear, in fact, to be curves running from the (0, 0) point to the (1, 1) point; they certainly are not single points in the unit square.

With fixed stimulus conditions, any experimental variable that generates an iso-sensitivity curve I shall call a *contingency variable*; examples are the instructions to the subject, the presentation probability of the stimulus, the information feedback to the subject about the accuracy of his performance, and payoffs. Figure 1 gives an example of the data obtained when payoffs are varied. It is not certain whether the several contingency variables all generate the same iso-sensitivity curve—most of the theories suggest that they do and certainly to a first approximation they do. The important conclusion from a large amount of data is that these non-sensory variables seriously influence a subject's performance in experiments designed primarily to study sensory phenomena.

In what sense can this be called a revolution when, after all, psychophysicists have long known that contingency variables affect behaviour? Certainly these effects are implicitly acknowledged when an experimenter, who is attempting to determine a threshold, introduces catch trials and then reprimands the

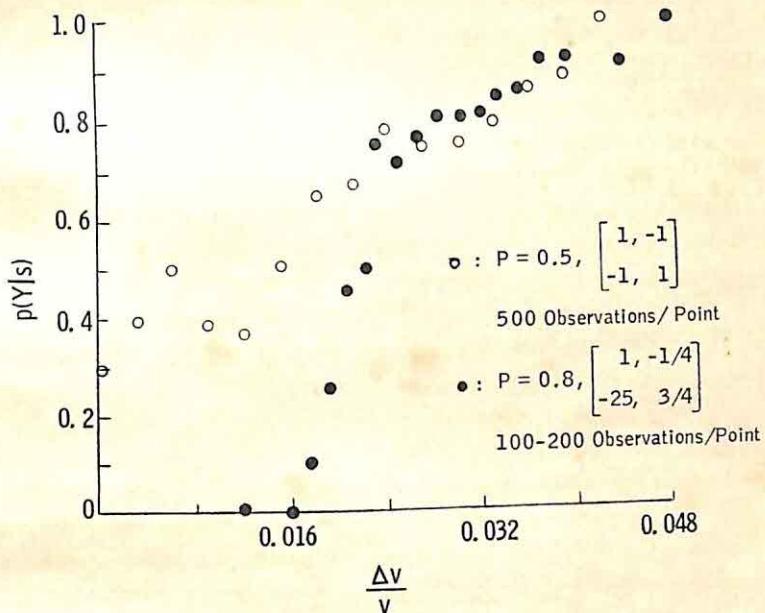


FIGURE 2. Psychometric functions (subject 6) reported by Norman (1962) for energy increments in a 1000 c.p.s. pure tone with the increment size varying from 0% to 4.8%. The payoff matrices and presentation probabilities are shown. Each point is estimated from 100 to 500 observations.

subject for not paying attention when he makes too many false positive responses. Such negative experiences usually have the desired effect of reducing the false alarms to an 'acceptable' level, and the experimenter is pleased with his skill at getting the subject to pay attention to the task; later, when reporting the data, he feels content to act as if the only relevant variables in the experiment were the physical parameters of the stimuli. Thus, for example, he speaks of *the* psychometric function, i.e., the function relating $p(Y|s)$ to a physical measure of s , as

if it were unique and quite independent of the reprimand. He seems to believe that the unreported, but low, false alarm rate was obtained only at the price of increased attention which permitted the unique psychometric function to shine through clearly, not beclouded by the subject's daydreaming. Perhaps this view is tenable, but one cannot help but be uneasy when one tries to decide just how severe to make the reprimand and just how often to administer it in order to arrive at the desired function. The difficulty becomes clear when we replace occasional reprimands by trial-by-trial payoffs and information feedback. One then generates families of psychometric functions by using different payoff matrices. Figure 2 shows two psychometric functions obtained from the same subject under identical stimulus conditions. The only difference is the payoff matrix and the presentation probability used. Which, if either, is *the* true psychometric function? Note that the one more like the usually reported psychometric funtions arises from a highly asymmetric payoff matrix.

Accompanying, and in some cases leading, this acceptance of the ubiquitousness and significance of contingency variables in psychophysics has been the development of new theories that included provision for their effects. In general, the theories are of the following form: the conditional probability of response r to stimulus s , $p(r|s)$, is some specified function of two classes of parameters, one of which is altered when and only when the stimulus conditions are altered and the other of which is altered when the contingency conditions are altered. For brevity, I shall refer to the former as *stimulus parameters* of the model and to the latter as *response-bias parameters*. Thus, an iso-sensitivity curve is generated by changes in a response-bias parameter with the stimulus parameters fixed, and the curve is unique if and only if there is a single response-bias parameter. A particular psychometric function is generated by changes in the stimulus parameters with the response-bias parameters held fixed.

Before turning to the main topic, let me point out that this new point of view can have appreciable substantive consequences. For example, several studies (Goetzl, Ahokas, & Payne, 1950; Hammer, 1951; Yensen, 1959) on taste thresholds for sucrose have been interpreted to mean that the threshold is higher following a meal than preceding it and, hence, that the ingestion of food affects sensitivity. Recent work by Dr. Mary Moore and her colleagues (Moore, Linker, & Purcell, 1964) in which the above notions were used in the analysis of the data, suggests that, on the contrary, the sensitivity is fixed but that the subject may exhibit a greater tendency to say 'yes' before eating than after eating. These results are consistent with the earlier ones in showing that $p(Y|s)$ is often reduced by eating, but in addition they show that the false alarm probability $p(Y|n)$ is also reduced by eating. The two changes are so coordinated that they appear to arise from a shift in bias, not a modification in sensitivity.

II. STATIC MODELS

As in most fields, the first psychophysical models to be examined were static in character. Although there are quite a number of models that differ in

many details, I believe that they all fall into one of two quite general classes. In one class, which may be called the *internal state models* and which includes, for example, Thurstone's law of comparative judgement, signal detectability theory, and the low and multi-state threshold models, the stimulus s is assumed to activate in a probabilistic fashion an internal state x which in turn leads via some decision mechanism to the response r . The overall response probability on trial m is assumed to arise from these two processes acting independently, thus leading to the expression

$$p_m(r|s) = \int p_m(r|x)p(x|s)dx, \quad (1)$$

where the integral is just that if there is a continuum of states or it is a sum if they are discrete. In the truly static models none of the probabilities depend upon the trial number m ; however, I have included m explicitly so that it will be clear what is assumed to vary when we turn to dynamic models. The conditional probability $p(x|s)$ represents the sensory mechanism and $p_m(r|x)$ the decision mechanism. In specific models assumptions are introduced about these probabilities and the free parameters thus introduced are called stimulus parameters in the first case and response-bias parameters in the second case.

In signal detectability theory (Green, 1960; Swets, 1961; Tanner & Swets, 1954), for example, $p(x|s)$ is assumed to be a normal density function and in a two-response situation $p_m(r|x)$ is assumed to have the form:

$$p_m(r|x) = \begin{cases} 1 & \text{if } x \geq c_m, \\ 0 & \text{if } x < c_m, \end{cases} \quad (2)$$

where c_m , which is usually called a cut-point or criterion, is the response-bias parameter. If one supposes that the subject chooses c_m so as to maximize his expected payoff, then c_m is a simple monotonic function of the experimental parameter

$$\beta = \left(\frac{1-P}{P} \right) \left(\frac{o_{22} - o_{21}}{o_{11} - o_{12}} \right), \quad (3)$$

where P is the probability that stimulus s_1 is presented and the payoff matrix is

$$\begin{array}{c} r_1 \quad r_2 \\ \begin{matrix} s_1 & \begin{bmatrix} o_{11} & o_{12} \\ o_{21} & o_{22} \end{bmatrix} \\ s_2 & \end{matrix} \end{array}. \quad (4)$$

It is known (Green, 1960) that the assumption that the expected payoff is maximized combined with the signal detectability model is not correct in detail. However, for many purposes having to do with the sensory aspects of the model, the dependence of c_m on the payoffs and presentation probabilities does not matter and so this overall failure has not caused much concern.

In the discrete (or threshold) models (Luce, 1963; Norman, 1962, 1963) the internal states form a discrete (finite or countably infinite) set. Let x_i denote the i th state. In some models certain of the parameters $p(r|x_i)$ are assumed to be

0 or 1 and the rest are free. Again, if we assume that the subject maximizes the expected payoff, the overall model is wrong if, as the data suggest, only a few states play any significant role. The reason is that the response probabilities are linear functions of the response bias parameters and so the maximum occurs with some of them equal to 0 and the rest equal to 1; this means that the iso-sensitivity curve is confined to a discrete set of points in the plane, in number one less than the number of states. Thus, for example, with four states the iso-sensitivity 'curve' would consist of only three points, which is not consistent with the data shown in Fig. 1.

The other class of models, specific examples of which have been discussed by a number of authors including Clarke (1957), Luce (1959), Shepard (1957), and Shipley (1960, 1961), have been called *choice models*. These models attempt to capture the belief of many psychologists that at least three independent processes combine to generate the responses observed in psychophysical experiments. First, the presentation of a stimulus results in a subjective magnitude that, in general, is non-linearly related to the usual physical measure of the intensity of the stimulus. Such a function is sometimes called a *psychophysical function*. Second, the subject exhibits some degree of confusion or generalization either between stimuli or between responses or both. Third, for a variety of reasons, a subject may exhibit a bias toward using one of the permitted responses as compared with another response. Let $\psi(s)$ denote the subjective magnitude of stimulus s , so ψ is the psychophysical function. Let $\zeta(t, r)$ be proportional to the amount of generalization from a subjective magnitude t to response r —in the model to be stated it is immaterial whether we suppose the generalization occurs at the level of responses or of stimuli. Finally, let $b_m(r)$ be proportional to the bias toward using response r on trial m . It has been assumed that ψ and ζ are fixed functions independent of the trial number, for they are assumed to describe two facets of the sensory apparatus, and that b is trial dependent, for it is assumed to represent the decision mechanism. For reasons that are too lengthy to go into in this paper, the following relation between the conditional response probability and these three functions has been postulated:

$$p_m(r|s) = K_m \zeta[\psi(s), r] b_m(r), \quad (5)$$

where K_m is simply the normalizing factor $1/\int \zeta[\psi(s), x] b_m(x) dx$. (For more details, see Ch. 5 of Luce, Bush, and Galanter, 1963.)

The model formulated in eqn. 5 is similar to the internal state model, eqn. 1, in the sense that it too postulates unobservable internal effects that are combined in a prescribed manner to generate the response probability. The two classes of models, although similar to this degree, are treated separately because so far as I know the choice models generally cannot be recast in terms of internal states satisfying eqn. 1.

Both the internal state and choice models involve unspecified and unobservable functions. In the former they are $p(x|s)$ and $p_m(r|x)$, and in the latter they are $\psi(s)$, $\zeta(t, r)$, and $b_m(r)$. Thus, a test of either model is possible only if

additional assumptions are made about the form of the unknown functions so as to reduce appreciably the amount of freedom involved, or if experiments are designed so that one or more of the functions play no effective role, or if the data from a number of closely related experiments are analysed simultaneously. Usually, some mixture of all three techniques is used. For example, with a few stimuli and responses, ψ is generally assumed to be the experimenter prescribed correspondence between stimuli and responses, ζ is assumed to satisfy certain constraints such as symmetry, i.e., $\zeta(t, r) = \zeta(r, t)$, the bias b is assumed to be constant over responses for certain symmetric experimental designs, and the stimulus parameters are assumed to be the same in going from a Yes-No to a forced-choice design.

As with the internal state models, auxiliary assumptions are needed to describe how the subject adjusts the bias parameter b with changes in presentation probabilities and payoffs. Once again, the evidence suggests that the choice model is incorrect when combined with the assumption that the expected payoff is maximized.

III. DECISION-MAKING VIEWED AS LEARNING

Although the evidence is far from conclusive as to where these psychophysical models go wrong, one particularly suspect feature is the assumption that the expected payoff (or even expected utility) is maximized. Unfortunately, the fact that the axiomatic structure of utility maximization is known is of little help in this problem. It is extremely difficult to design experiments adequate to test this hypothesis directly within the context of psychophysics. Among the problems, not the least is the fact that the behaviour appears to be probabilistic whereas the relevant utility models are algebraic.

The only alternative decision process that has been seriously investigated, and that for only a few years, is the assumption that subjects adapt and adjust their response biases in the light of their recent experience in the experiment; that is, they learn. Surely such an assumption is *a priori* reasonable, and any psychophysicist is aware that it or some other sort of systematic change occurs over trials before the behaviour settles down to the 'asymptotic' values that are usually reported. So the question is whether we can effectively couple a stochastic learning process with a model for the sensory process in such a way that its asymptotic predictions describe behaviour in detection, recognition, and discrimination experiments.

Quite a number of possibilities are available, only a few of which have been examined so far. Specifically, we can distinguish three types of sensory processes according to whether they postulate no internal state, a discrete set of states, or a continuum of them. Second, we have several alternative learning models, including the linear operators, certain non-linear commutative operators, and the family of Markov chain models that arise from stimulus sampling theory. Finally, there is the question of whether the learning applies to the sensory

mechanism, the decision mechanism, or both. So at a minimum, there are $3^3 = 27$ possible models. I shall discuss only a few of these, for only a few have been worked out.

IV. SINGLE-PROCESS LEARNING MODELS

Let us examine first a class of single-process learning models in which the sensory mechanism is distinctly subordinate to the learning. Until recently most applications of mathematical learning theories have been to so-called simple learning situations in which the same stimulus condition obtains on every trial; however, serious work has now begun on models for what are usually called discrimination experiments in learning. It should be noted that their formal analogues in psychophysics are detection and recognition experiments, not psychophysical discrimination experiments. The main difference between the psychophysical and learning versions of these formally identical experiments is that in the psychophysical setting the experimenter-determined connection between stimuli and responses is a 'natural' one that is known to the subject, but the stimuli are difficult for him to identify; whereas, in the learning context, the connection that the experimenter has established between the stimuli and responses may not be particularly natural and, more important, initially is not known to the subject, but the stimuli, once they are attended to, are easy to identify. In both classes of experiments, the subject's association of a particular response to a particular stimulus is usually measured by the estimated conditional response probabilities. The simplest learning model to consider is the one in which these conditional response probabilities increase or decrease systematically as a function of what has just occurred to the subject in terms of stimulation, responses, and outcomes. These are the single-process models.

As an illustration, consider the simple case of two responses and two stimuli, with response r_1 correct for stimulus s_1 and response r_2 correct for stimulus s_2 . Since the response probabilities have to sum to 1, it is sufficient to consider what happens to the probability of r_1 when s_1 is presented and also when s_2 is presented. If an s_1 occurs on trial m —this the subject only finds out after his response—then it is plausible to suppose that $p(r_1|s_1)$ is increased, perhaps by a linear operator of the form

$$p_{m+1}(r_1|s_1) = (1 - \theta_1)p_m(r_1|s_1) + \theta_1. \quad (6)$$

If, however, the stimuli are difficult to identify, then there should be some generalization from stimulus 1 to stimulus 2 and so $p(r_1|s_2)$ should also increase, but by a lesser amount. If we again assume that a linear operator describes the change we can write it as

$$p_{m+1}(r_1|s_2) = (1 - \eta_2\theta_1)p_m(r_1|s_1) + \eta_2\theta_1, \quad (7)$$

where we represent the learning rate parameter as a proportion η_2 of the learning rate parameter for stimulus 1, so $0 \leq \eta_2 \leq 1$. In like manner, if s_2 occurs, then $p(r_2|s_2)$ increases linearly by a direct effect and $p(r_2|s_1)$ increases linearly by

generalization or, what is the same thing, $p(r_1|s_2)$ decreases directly and $p(r_1|s_1)$ indirectly. If we suppose s_1 and s_2 are presented according to a simple random schedule with $P = \Pr\{s_1\}$ and $1 - P = \Pr\{s_2\}$, then it is not difficult to show (Bush, Luce, & Rose, 1963) that the asymptotic expected response probabilities are of the form

$$P_\infty(r_1|s_1) = \frac{1}{1 + \eta_1 b} \text{ and } P_\infty(r_1|s_2) = \frac{\eta_2}{\eta_2 + b}, \quad (8)$$

where

$$b = \left(\frac{1-P}{P}\right) \frac{\theta_2}{\theta_1}.$$

Since we interpret η_1 and η_2 as representing stimulus generalization, they are the stimulus parameters of the model. The parameter b is interpreted as a response bias parameter which depends explicitly upon the presentation probability and implicitly upon the instructions and payoffs through the learning rate parameters.

These asymptotic predictions are identical to those given by the static choice model mentioned earlier (eqn. 5), and so this linear learning model can be treated as a justification of that static model for two stimuli and two responses. This learning analysis can be generalized readily to the case of k responses to k stimulus presentation and only slightly less readily to the case of 2 responses to k stimulus presentations, which is the design usually used in psychophysical discrimination experiments, and again close relationships to the choice model are found. Certain interesting subtleties arise when the number of responses differs from the number of stimulus presentations which suggest, for example, that data obtained using the method of constant stimuli are much more complicated to analyse than those obtained using pair comparisons (see Bush, Luce, & Rose, 1963).

If one looks at the iso-sensitivity curve that is predicted when η_1 and η_2 are held fixed and b is varied from 0 to ∞ , it is seen to be a symmetric curve about the diagonal from $(0, 1)$ to $(1, 0)$, as is the curve predicted by the signal detectability model with equal variances. For human Yes-No detection data this is clearly wrong and for two-alternative forced-choice detection data, which do appear to be symmetric, it may very well be of the wrong shape. For human recognition experiments in which perfectly detectable stimuli are to be identified it has not yet been shown to be wrong. For animal experiments it is unclear whether or not one can get asymptotic behaviour which lies interior to the unit square. If the stimuli are perfectly discriminable, then asymptotically the animals usually learn and end up at $(0, 1)$; if the stimuli are not perfectly discriminable or if the animals do not learn, they frequently end up at $(0, 0)$ or $(1, 1)$ which are position habits or, in present terms, pure response biases. The recent data of Hack (1963), although variable and surely not asymptotic, suggest that perhaps intermediate behaviour is possible. Much more work on the psychophysics of animals is needed before we can be sure what the qualitative facts are.

If one sets up the analogous learning process with stimulus generalization using the non-linear beta model operators (Luce, 1959) of the following form

$$v_{m+1} = \beta v_m \text{ and } u_{m+1} = \beta^* u_m, \quad (9)$$

where

$$p_m(r_1|s_1) = \frac{v_m}{1+v_m} \text{ and } p_m(r_1|s_2) = \frac{u_m}{1+u_m},$$

then Bush (1964) has shown that at asymptote either $P_\infty(r_1|s_1) = 1$ or $P_\infty(r_1|s_2) = 0$ or both provided that the parameters are chosen to produce changes paralleling those of the above linear model. Such a model is clearly wrong for human data, but perhaps it may describe how animals perform.

V. Two-PROCESS MODELS

Most students of psychophysics believe that the single-process models of the type just described are too simple to account adequately for what underlies the subject's response behaviour. At the very least they feel that the perceptual and decision stages of the process must be dealt with separately and that the two classes of mechanisms are somewhat different. Perhaps the most prevalent view is that the sensory process is static in the sense that it is totally unaffected by the contingency variables of the experiment and that all of the learning takes place in a decision process that converts the internal sensory states into responses. Thus, for example, in signal detectability theory one assumes that the normal distributions over the continuum of states depend only upon the stimulation and the subject, but not on the payoffs or instructions. The contingency variables are assumed to affect only the subject's choice of the criterion c_m . Thus, the learning process is assumed to describe how the subject adjusts c_m from trial to trial.

Relatively little has yet been done in the learning literature that seems applicable to continuous processes of this type. The only suggestion that I am aware of is due to Kac (1962) who postulated the following random walk for the criterion. Let c_m denote the criterion and x_m the internal state on trial m , then Kac assumed that

$$c_{m+1} = c_m + \begin{cases} \Delta & \text{if } s_2 \text{ is presented and } x_m > c_m \\ -\Delta' & \text{if } s_1 \text{ is presented and } x_m < c_m \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

An exact solution for the asymptote of this process is not known, but by making the approximation $F(a+x) \approx F(a) + xF'(a)$, where F is a cumulative normal distribution, Kac showed that the asymptotic expected criterion, c_∞ , is the solution to

$$\Delta'P \int_{-\infty}^{c_\infty} f_{s_1}(x)dx = \Delta(1-P) \int_{-\infty}^{c_\infty} f_{s_2}(x)dx,$$

where f_{s_1} and f_{s_2} are the normal density functions corresponding to stimuli s_1 and s_2 . No tests of this model have been reported. Also, it is not clear whether

the first two terms of the Taylor series expansion of F is an acceptable approximation.

For models with a discrete set of states, it is relatively easy to introduce a learning process. In essence, one proceeds as follows. If state x_i occurred on trial m and was due to stimulus s_k , then the probability of making the response r_k that is appropriate to s_k when x_i occurs, $p(r_k|x_i)$, is increased. The probabilities of making any other response to x_i are decreased, and the probabilities having to do with the other internal states remain unchanged. If one assumes linear operators—which is the only case that has been worked out (Luce, 1963; Norman, 1962, 1963)—then the asymptotic expected biases for the two-response and two-stimulus case are of the form

$$p_\infty(r_1 | x_i) = \frac{p(x_i | s_1)}{p(x_i | s_1) + p(x_i | s_2)b}, \quad (12)$$

where

$$b = \left(\frac{1-P}{P} \right) \frac{\theta_2}{\theta_1}.$$

The response probabilities are given by

$$p_\infty(r_1 | s_k) = \sum_{i=0}^{\infty} p_\infty(r_1 | x_i) p(x_i | s_k), \quad (13)$$

where the $p(x_i | s_k)$ are the stimulus parameters. For the Yes-No detection experiment, this leads to predictions of psychometric functions of the form shown

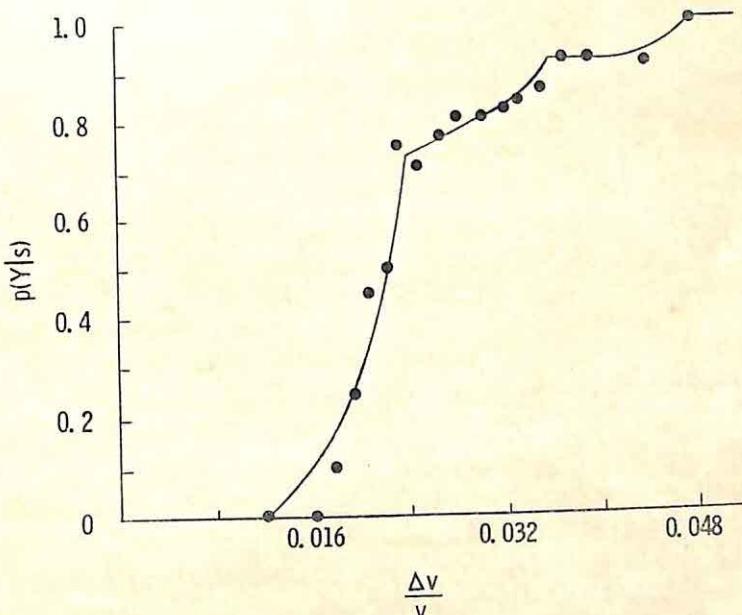


FIGURE 3. The multi-stage threshold model psychometric function compared with the data of Fig. 2. The parameters of this curve were chosen to fit all of the data for subject 6, not just the data shown.

in Fig. 3. Data which suggest that such a complicated structure may exist have been obtained by Larkin and Norman (1963) and Norman (1962, 1963); a sample is shown in the figure.

The iso-sensitivity curves for the multi-state threshold model of the Yes-No detection experiment, using parameter values appropriate to the data (which, among other things, means that only the first four states contribute significantly) are asymmetric curves, not unlike those of unequal-variance signal detectability theory but with a somewhat sharper corner. A typical curve, fitted to data (Norman, 1962), along with the bilinear prediction of the two-state model (Luce, 1963), is shown in Fig. 4.

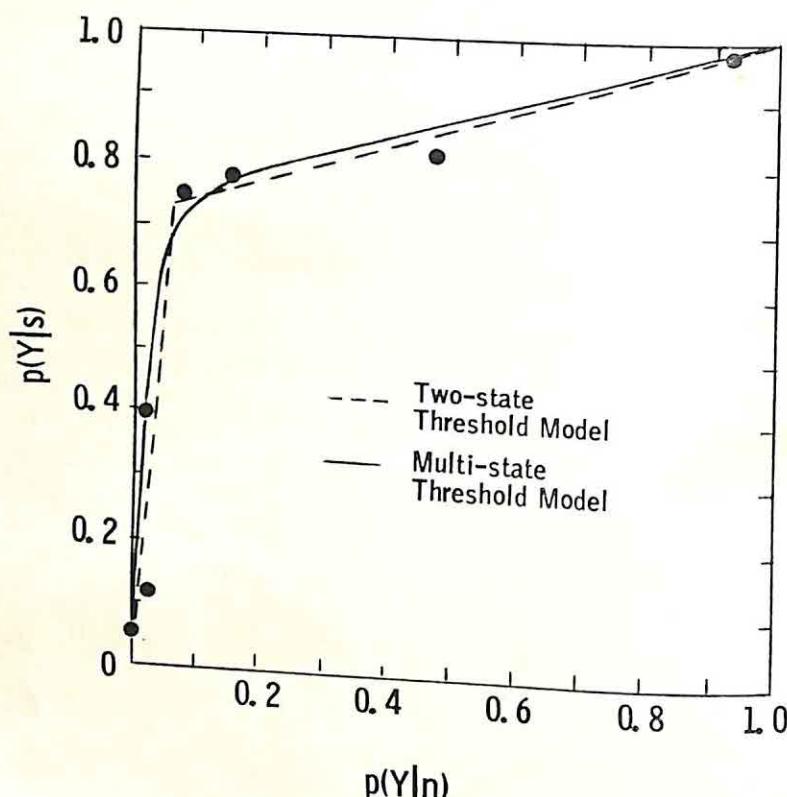


FIGURE 4. The multi-stage and two-state threshold model iso-sensitivity curves compared with the data on Fig. 1. The parameters of the multi-stage model were chosen to fit all of the data for subject 6, not just the data shown.

The only other two-process learning model for psychophysical experiments of which I am aware is the recent one proposed by Atkinson (1963). It is rather too complicated to describe in detail here. Suffice it to say that it differs not only in the way that stimulus sampling models all differ from operator models, but also in its fundamental conception of what is modified by the learning. Atkinson postulates that the subject's sensitivity to the stimulus is varied systematically as

a result of experience. Like all of the models discussed above, his has a fair number of free parameters to be estimated from the data, and this appears to give it adequate freedom to account for most of the functions that traditionally have been studied. In addition, he has focused considerable attention on the sequential properties of the data. It is too early to evaluate the adequacy of these results, especially since replications of nearly the same experiments have not yielded entirely consistent data.

VI. CONCLUDING REMARKS

The task of trying to incorporate learning ideas into formal models for psychophysical experiments has barely begun, and there are innumerable directions the research might take. I shall conclude by listing three of the directions that I feel would be most useful in the near future.

1. A more complete learning analysis is needed to describe the adjustment of the criterion when there is a continuum of states, as in detection theory. Since various learning theorists have, for independent reasons, begun to work on continuous learning problems, perhaps some transfer of ideas will be possible.

2. In the two-process models, there is a choice whether the learning applies to the sensory or decision process, and so it would be very helpful indeed if one were to be able to derive fairly sharp differential predictions that could be used to decide between the two assumptions.

3. In several of the models, the response-bias parameter is expressed as a function of the presentation probability and the learning rate parameters. Since we know that payoffs affect the biases, it must be that the learning rate parameters depend in some fashion on the payoffs. They may also depend upon other things, including the presentation probabilities. A theory for their dependence upon, at least, the payoffs would reduce considerably the number of free parameters and might very well lead to some interesting new predictions.

REFERENCES

- ATKINSON, R. C. (1963). A variable sensitivity theory of signal detection. *Psychol. Rev.*, **70**, 91-106.
- BUSH, R. R. (1964). Identification learning. In R. D. Luce, R. R. Bush, E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 3. New York: Wiley. (In press.)
- BUSH, R. R., LUCE, R. D., and ROSE, R. M. (1963). Learning models for psychophysics. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press. (In press.)
- CLARKE, F. R. (1957). Constant-ratio rule for confusion matrices in speech communication. *J. Acoust. Soc. Amer.*, **29**, 715-720.
- GOETZL, F., AHOKAS, ANN, and PAYNE, JEAN (1950). Occurrence in normal individuals of diurnal variations in the sense of taste for sucrose. *J. appl. Physiol.*, **2**, 619-626.
- GREEN, D. M. (1960). Psychoacoustics and detection theory. *J. Acoust. Soc. Amer.*, **32**, 1189-1203. Reprinted in R. D. Luce, R. R. Bush and E. Galanter (Eds.), *Readings in mathematical psychology*, Vol. 1. New York: Wiley, 1963. Pp. 41-66.
- HACK, M. H. (1963). Signal detection in the rat. *Science*, **139**, 758-760.

- HAMMER, F. J. (1951). The relation of odor, taste, and flicker-fusion thresholds to food intake. *J. comp. physiol. Psychol.*, **44**, 403-411.
- KAC, M. (1962). A note on learning signal detection. *I.R.E. Trans. on Info. theory*, **IT-8**, 126-128.
- LARKIN, W. D., and NORMAN, D. A. (1963). An extension and experimental analysis of the neutral quantum theory. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press. (In press.)
- LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- LUCE, R. D. (1963). A threshold theory for simple detection experiments. *Psychol. Rev.*, **70**, 61-79.
- LUCE, R. D., BUSH, R. R., and GALANTER, E. (1963). *Handbook of mathematical psychology*, Vol. 1. New York: Wiley.
- MOORE, MARY E., LINKER, E., and PURCELL, MARGUERITE (1964). Taste sensitivity after eating: a signal detection approach. Submitted for publication.
- NORMAN, D. A. (1962). *Sensory thresholds and response biases in detection experiments: a theoretical and experimental analysis*. Ph.D. thesis, Dept. of Psychol., U. of Penna.
- NORMAN, D. A. (1963). Sensory thresholds and response bias. *J. Acoust. Soc. Amer.*, **35**, 1432-1441.
- SHEPARD, R. N. (1957). Stimulus and response generalization: a stochastic model relating to distance in a psychological space. *Psychometrika*, **22**, 325-345.
- SHIPLEY, ELIZABETH F. (1960). A model for detection and recognition with signal uncertainty. *Psychometrika*, **25**, 273-289.
- SHIPLEY, ELIZABETH F. (1961). *Detection and recognition with uncertainty: an experiment and a revised model*. Ph.D. thesis, Dept. of Psychol., U. of Penna.
- SWETS, J. A. (1961). Detection theory and psychophysics: a review. *Psychometrika*, **26**, 49-63.
- TANNER, W. P., JR., and SWETS, J. A. (1954). A decision making theory of visual detection. *Psychol. Rev.*, **61**, 401-409.
- YENSEN, R. (1959). Some factors affecting taste sensitivity in man: food intake and time of day. *Quart. J. exp. Psychol.*, **11**, 211-229.

THE EFFECT OF ONE STIMULUS ON THE THRESHOLD FOR
ANOTHER : AN APPLICATION OF SIGNAL DETECTABILITY
THEORY¹

By MICHEL TREISMAN

Institute of Experimental Psychology, Oxford.

If an accessory stimulus regularly precedes a critical stimulus at a fixed inter-stimulus interval then the shorter the interval the lower the threshold for the critical stimulus. Three hypotheses which might explain this are discussed, and two experiments described. Experiment 1 shows that varying the intensity of an accessory stimulus does not affect the extent of the threshold fall it produces. Experiment 2 confirmed a prediction that randomizing inter-stimulus intervals over a small range would allow the threshold to fall as the mean inter-stimulus interval decreased at long mean intervals but not at short ones. Taken with earlier results, these experiments indicate that the subject lowers his threshold during a range of time about the end of the inter-stimulus interval in which he expects that the stimulus may occur. It is also shown that the extent of the fall in threshold is inversely proportional to the length of this range of expectation, and that the latter appears to be directly proportional to the duration of the inter-stimulus interval.

This variation in threshold level is considered in terms of the model of the threshold provided by signal detectability theory. It is shown that the threshold falls are due to changes in the criterion computed by the subject, rather than in the sensory noise. This model is also successfully used to predict the extent of the shifts in detection rate that occur as the inter-stimulus interval alters, on the assumption that subjects attempt to keep their false positive rates constant. Finally it is suggested that in computing the appropriate criterion for each inter-stimulus interval the subject uses a simple approximate method rather than the more complex exact method.

I. INTRODUCTION

If an accessory stimulus is regularly given at a fixed interval before a critical stimulus it alters the threshold for the latter. Figure 1 shows the results of five experiments performed by Howarth & Treisman (1958). It can be seen that as the interval between the accessory and critical stimuli is reduced from 9 to 1 seconds the threshold falls, slowly at first and then more rapidly. Curves 1, 2, 3 and 4 show the absolute threshold for the electric phosphene, measured by the method of limits, and curve 7 gives the differential threshold for the intensity of a 500 c.p.s. tone. A light-flash regularly preceded the critical stimulus for curves 1 and 2, a bell for curves 3 and 4, and a light for curve 7.

¹ I would like to thank Professor R. C. Oldfield for the provision of research facilities, and Mrs. J. Clarke for assistance with computing. Part of the work described here was done while I was in receipt of an MRC grant.

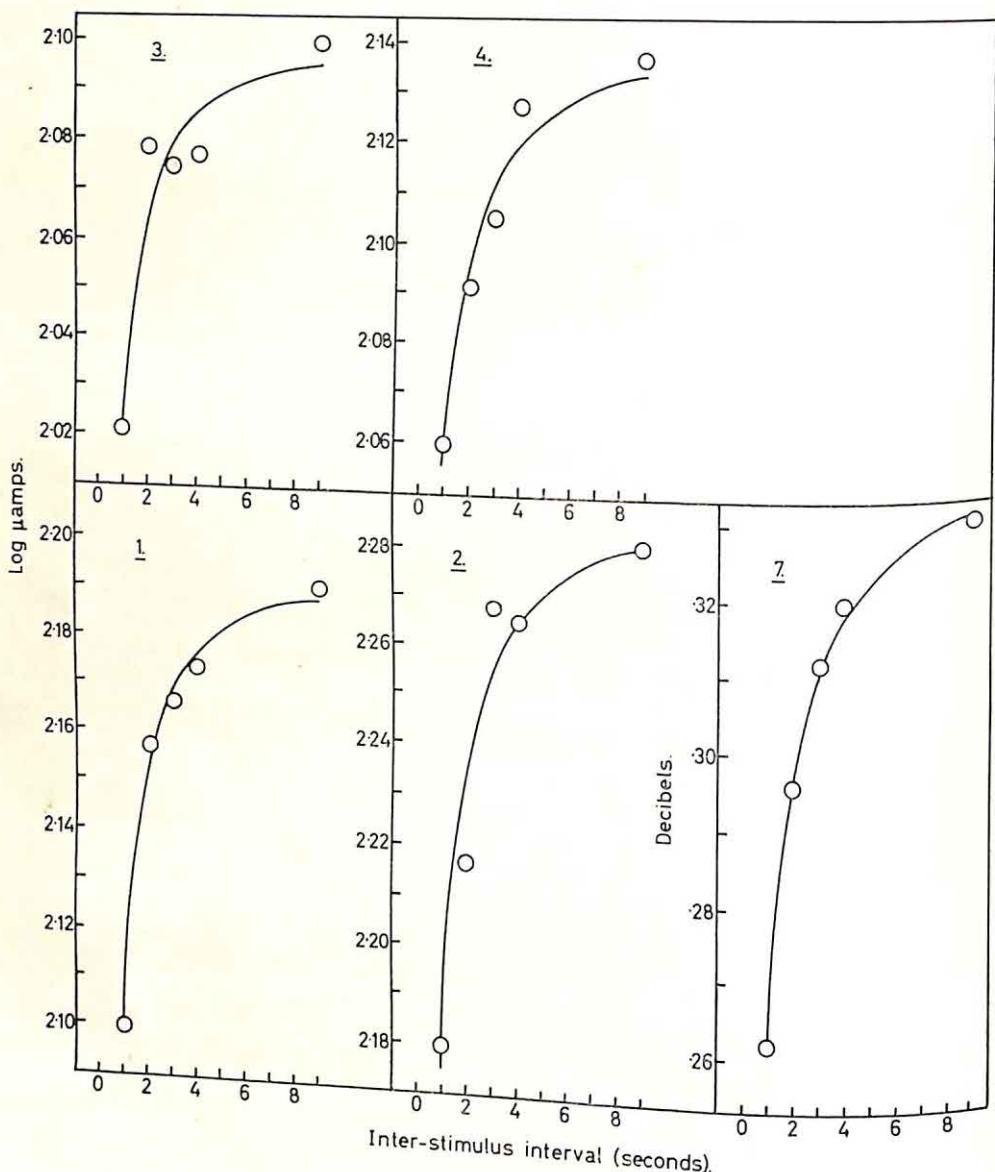


FIGURE 1. Data taken from Howarth & Treisman (1958). Curves 1, 2, 3 and 4 show absolute thresholds for the electric phosphene, measured in $\log \mu$ amps, and curve 7 shows differential thresholds for increases in the intensity of a 500 c.p.s. tone, in db. Each threshold was determined by the method of limits, the critical stimulus being given at a fixed interval after an accessory stimulus; the latter was a flash of light for curves 1 and 2, the ringing of an electric bell for curves 3 and 4, and the onset of a neon light for curve 7. The fixed intervals used were 1, 2, 3, 4 or 9 seconds. Theoretical curves have been fitted to these data, as described in Section V. Their equations are: (1) $2.20-L=0.07/(T-0.15)$; (2) $2.29-L=0.11/(T-0.05)$; (3) $2.10-L=0.08/(T-0.15)$; (4) $2.14-L=0.10/(T+0.15)$; (7) $0.34-L=0.10/(T+0.25)$.

Thus the effect does not depend on either the critical or accessory stimulus being in a particular modality. Treisman & Howarth (1959) showed that the false positive rate does not increase as the threshold for the critical stimulus falls, so that this fall cannot be attributed to an increase in guessing.

These experimental findings pose two problems: (1) How does the accessory stimulus have this effect and, in particular, how does the length of the inter-stimulus interval determine the extent of the fall in threshold produced by the accessory stimulus? (2) What mechanism of sensory discrimination allows a fall in threshold to occur without an increase in the false positive rate? The first part of this paper is directed to the first question. Different hypotheses to account for the effect of the accessory stimulus will be discussed, and two experiments will be presented which, taken with earlier results, allow us to prefer one of these hypotheses. The second part of the paper will develop this hypothesis in relation to the model of the threshold given by signal detectability theory, and will show that the fuller account which this leads to allows predictions to be derived which agree reasonably well with experimental results.

II. ALTERNATIVE HYPOTHESES

To explain the form of the relation between inter-stimulus interval and threshold level two hypotheses have been put forward (Howarth & Treisman, 1958, 1961; Treisman & Howarth, 1959): (a) The accessory stimulus causes an immediate arousal or alerting response, one effect of which is a rapid increase in sensitivity to incoming stimuli. The threshold curves obtained reflect the time-course of the decay of the physiological change underlying this. This will be referred to as the 'arousal' hypothesis. (b) The accessory stimulus acts as a warning or temporal reference point, allowing the subject to use his knowledge of the inter-stimulus interval to anticipate the arrival of the critical stimulus, and to lower his threshold when he expects it. Since he cannot determine the end of a time interval exactly, he will expect the stimulus over a range of time about the end of the interval. Thus if the inter-stimulus interval is 3 seconds, the variability of the subject's measure of time might result in his expecting the stimulus to occur at times between 2.5 and 3.5 seconds after the accessory stimulus, but not earlier or later than this period, and to lower his threshold only during this one second range of intervals. The length of this 'range of expectation' would be approximately proportional to the inter-stimulus interval; it is assumed that the threshold is lowered to an extent inversely related to the length of this range (Howarth & Treisman, 1958). This will be referred to as the 'range of expectation' hypothesis. (c) To these alternatives a third hypothesis can be added. This also assumes that the accessory stimulus functions by providing information rather than by producing arousal, but it supposes that the subject lowers his threshold immediately he receives the accessory stimulus, rather than waiting for and anticipating the end of the inter-stimulus interval. The threshold is lowered to a fixed level at which the subject maintains it throughout the time that he waits for the stimulus, or until he concludes

that the trial is over. This waiting time would be about the same length or a little longer than the inter-stimulus interval. The subject uses his knowledge of the length of the waiting time to determine the extent of the fall in threshold instituted when the accessory stimulus is received: the threshold fall is inversely related to the anticipated waiting time. This will be referred to as the 'waiting time' hypothesis.

These hypotheses were tested by an experiment in which the subjects' information about the length of the inter-stimulus intervals was reduced by varying them at random, on successive trials, between 0.5 and 5.5 seconds. Such a loss of temporal information should not affect a 'reflex' physiological change, but it should make the subject unable to anticipate the critical stimulus successfully. It was found that the relation between inter-stimulus interval and threshold previously found was no longer shown (Howarth & Treisman, 1958).

Though this tells against the arousal hypothesis it is not conclusive. It might be that the 'arousal response' evoked by the accessory stimulus normally habituates (Pavlov, 1927; Humphrey, 1933), but that this is prevented when there is the 'reinforcement' of having the critical stimulus come at about the expected time. In an experiment in which no temporal information is given habituation might proceed unchecked, the relation between inter-stimulus interval and threshold disappearing. Temporal information could, then, be needed to maintain the 'arousal response', and might not be used to anticipate when the stimulus will arrive. We may also note that the result of this experiment is compatible with both the range of expectation and the waiting time hypotheses. In the latter case randomizing stimulus intervals might have resulted in the waiting time being the same on all trials. Two experiments which test these hypotheses further will now be presented.

III. EXPERIMENT 1

The intensity of the stimulus may affect the amplitude of a reflex, the length of reaction time, the direction of attention, and the rate of habituation (Humphrey, 1933; Oldfield, 1937). We might, therefore, expect an increase in sensitivity due to arousal to be similarly affected. If, however, the accessory stimulus acts only as a temporal reference point its intensity should be irrelevant to its effect. This makes it of interest to see whether an effect of the intensity of the accessory stimulus can be shown when it is used to cause a fall in the threshold for a critical stimulus. To test this two accessory stimuli were used, one just above the threshold range and the other considerably stronger but not so intense as to be disturbing. Two inter-stimulus intervals, 0 and 1.5 seconds, were used, and the threshold for the critical stimulus was determined by the method of constant stimuli for each combination of accessory stimulus strength and inter-stimulus interval. The threshold was also determined without an accessory stimulus.

Apparatus

The subject sat alone in a moderately illuminated experimental room, 85 cm. from a white cardboard screen, 50 cm. \times 60 cm. in size. A 'pre-warning' stimulus was used to indicate the onset of each trial. This was provided by a neon bulb covered with a paper shade and placed at the right edge of the screen, 25 cm. from the fixation circle. The subject wore Brown type-K moving coil earphones through which he received a constant 500 c.p.s. 60-70 dB SL tone from a Muirhead audio-oscillator; the critical stimulus was a 40 msec. increase in the intensity of this tone. At the centre of the screen was a circular aperture 7 cm. in diameter which subtended 4° 40' of arc and was filled with a 0.2 Neutral Density Ilford filter. A fluorescent tube was contained in a lightproof box behind the fixation circle and some distance from it; when a 40 msec. pulse of current passed through it the light flash produced illuminated the circle diffusely. This constituted the accessory stimulus. Its intensity could be varied by placing neutral filters between the fluorescent tube and the screen. The brightness of the screen was 3.2 foot lamberts, the unilluminated fixation circle 0.40 foot lamberts. The 'weak' (W) light flash was 0.56 foot lamberts, and the 'strong' (S) flash 1.80 foot lamberts. To prevent the pre-warning stimulus providing any temporal information the interval between it and the accessory stimulus was randomized over a wide range.

The electronic timing circuits which were used to produce the random and fixed intervals, and the amplifier which allowed the intensity of the critical stimulus to be varied have been described elsewhere (Howarth & Treisman, 1958). Dekatron counters were used to record time intervals.

Procedure

At the beginning of each session the subject's threshold was estimated using the method of limits. The experimenter then selected four equally spaced intensities of the critical stimulus within the subject's threshold range. These were presented in series of 20 trials, each series consisting of 4 of each of the stimulus strengths, and 4 'blanks', in random order. A series was given under one of five combinations of accessory stimulus and inter-stimulus interval: W/0, W/1.5, S/0, S/1.5, and NAS (i.e., no accessory stimulus). Fifteen series, 3 under each combination, were presented during a session, in a random order different for each subject. Sessions lasted 1-2 hours. Five subjects (undergraduates and research students) each did one session.

Each trial began when the pre-warning neon bulb came on (it stayed on until the subject had responded). This was followed, after an interval varying at random between 2 and 5 seconds, by the accessory stimulus. The critical stimulus was given either simultaneously with or 1.5 seconds after the accessory stimulus (or replaced it: the NAS condition). The subject responded 'Yes' or 'No' by depressing one of two tapping keys. Reaction times were recorded, but subjects were not told that this was being done.

TABLE I. RESULTS OF EXPERIMENT 1

Conditions (see text for explanation)

<i>Response Measures</i>	NAS	W/0	S/0	W/1.5	S/1.5
Threshold (dB):	0.576	0.511	0.518	0.536	0.500
Standard Errors:	0.013	0.019	0.016	0.013	0.029
False Positives (per cent):	6.7	3.3	1.7	0.0	6.7
Reaction Times (seconds):	1.108	1.010	0.849	0.723	0.804
Standard Errors:	0.077	0.041	0.027	0.028	0.032

Results

The thresholds for each subject under each condition were determined by probit analysis (Finney, 1952). They are given in Table I. Each false positive rate is based on 60 'blanks'. The mean reaction times for 'Yes' responses are also given. They are derived from 647 positive responses.

Discussion

As previously found, the threshold falls when an accessory stimulus is given, and this is not attributable to a rise in the false positive rate. When compared with the NAS threshold all the other thresholds show significant falls ($p < 0.01$ for W/0 and S/0, $p < 0.02$ for S/1.5, $p < 0.05$ for W/1.5, on *t*-test). Despite this clear effect, at neither interval is there a significant difference between the thresholds obtained with the two accessory stimulus strengths. It might have been that the difference in the intensities of the W and S stimuli was not sufficient to cause different degrees of arousal. But if reaction time is accepted as an index of arousal then the subjects' response times provide evidence on this. When the accessory and critical stimuli are simultaneous, the W and NAS reaction times are not significantly different, but the S flash produces responses significantly more rapid than the W flash ($p < 0.001$) or the NAS condition ($p < 0.01$). Thus the two intensities of accessory stimulation used here have a differential effect on the speed of response, an effect readily ascribed to differences in the degree of arousal. The failure to find any similar effect on the thresholds does not exclude a 'reflex' action on sensitivity, but does make it less plausible.

These results suggest that the accessory stimulus may have two effects: an increase in arousal which can shorten the reaction time to a stimulus, this effect being greater if the inter-stimulus interval has some small positive value than if it is zero, and an effect on the threshold which is immediately apparent at the zero interval. The familiar observation that as the intensity of a stimulus increases the reaction time to it shortens is extended by these results to the case where it is the intensity of an accessory stimulus, not that of the stimulus evoking the response, that varies.

IV. EXPERIMENT 2

Although Experiment 1 provides further evidence against the arousal hypothesis it is not conclusive. We also need to distinguish experimentally between the 'waiting time' and 'range of expectation' hypotheses. In the next experiment a procedure which leads to different predictions for the arousal and 'waiting time' hypotheses on the one hand, and the 'range of expectation' hypothesis on the other, is used. If inter-stimulus intervals are randomized over a small range only, then there will still be a relation between the accessory stimulus and the time of arrival of the critical stimulus so that 'arousal'

'responses' should not habituate. It follows that in this case the arousal hypothesis will predict a fall in threshold as the mean inter-stimulus interval decreases similar to that found previously. On the 'waiting time' hypothesis the threshold is kept constant at a lowered level from the time of arrival of the accessory stimulus until the subject has detected or no longer expects the critical stimulus. If the inter-stimulus intervals vary over a small range the mean waiting time should be about as long as the longer intervals; therefore as the mean inter-stimulus interval decreases the threshold should fall. The 'range of expectation' hypothesis leads to a different prediction. When the mean inter-stimulus interval is sufficiently long the range of expectation will be considerably greater than the range over which the intervals are experimentally randomized, so that this randomization will have little effect. When, however, the inter-stimulus intervals are sufficiently short for the range of randomization to be somewhat larger than the range of expectation, then the frequency of positive responses to a given stimulus will become constant, and it will not increase as the mean inter-stimulus interval further decreases. (This will be true whether the subject continues to reduce his range of expectation as though there were no experimental variation of the inter-stimulus interval—the threshold in the range of expectation would fall but so would the proportion of stimuli falling within this range—or ceases to reduce it once it equals the range of randomization.) Thus this hypothesis predicts a decline in threshold as the mean inter-stimulus interval decreases, for longer intervals, but a plateau for the shorter intervals, where the other two hypotheses predict the steepest falls.

Inter-stimulus intervals ranging between 0 and 3 seconds were used in this experiment, since in the earlier work the largest falls in threshold occurred here. The range of expectation at the middle time interval was estimated by supposing it would be approximately equal to 5–6 times the differential threshold for temporal duration, ΔT , when the latter is defined as the standard deviation of productions or reproductions of a time interval, T . In an experiment described elsewhere the relation of ΔT to T was determined. This gave $\Delta T = 0.19$ seconds for $T = 1.5$ seconds, so a 1-second range of randomization was chosen (Treisman, 1963 a).

Apparatus

The subject sat facing a white cardboard screen with two neon bulbs at its centre. The upper one came on to give a pre-warning, and the accessory stimulus was provided by the lower one. They both stayed on until the subject had responded on each trial. Subjects responded verbally 'Yes' or 'No'; they were in telephonic communication with the experimenter. The remaining apparatus was as in Experiment 1. The critical stimulus was a 50 msec. increment in the intensity of the constant 500 c.p.s. tone.

Procedure

At the beginning of each session the subject's threshold was estimated by the method of limits, and two intensities of the critical stimulus, one 'weak' (W) and one 'strong'

(S), 0·1 db apart, were chosen near the threshold. In the course of each session 120 W, 120 S, and 12 'blank' stimuli were given. There were two Conditions: in the Constant Condition (CC) the stimuli were given in 12 series of equal length, each containing an equal number of W and S stimuli in random order. Two series were given at each of 6 inter-stimulus intervals: 0·25, 0·75, 1·25, 1·75, 2·25 and 2·75 seconds, in a random order different for each subject. In the Randomized Condition (RC) the stimuli were given in 6 randomly ordered series of equal length, 2 at each of 3 ranges of inter-stimulus intervals: 0·1 seconds, 1·2 seconds and 2·3 seconds. The six series were given in a counter-balanced order different for each subject. The randomizing circuits distributed the stimuli rectangularly over each 1 second range with half the W and half the S stimuli falling in each successive half-second, called here a 'sub-range'.

Each trial began with a 1-second pre-warning interval. Before each series of trials the subject was told the length of the inter-stimulus interval, or the range of intervals that would be employed. Sessions lasted about an hour. The subjects were 8 undergraduates. Each did 2 sessions, one under each condition, four with RC first and four with CC first.

Results

The percentages of positive responses to the 'strong' stimulus (P_S), the 'weak' stimulus (P_W), and both together (P_M), were calculated for each inter-stimulus interval or half-second 'sub-range'. They are given in Table II. The percentages of response for each subject were submitted to an angular transformation and an analysis of variance. Since no significant interaction between the effects of inter-stimulus intervals and stimulus intensities was found the mean percentages of positive response to both stimulus strengths, P_M , are plotted in Figure 2.

Discussion

The effect shown is that predicted by the range of expectation hypothesis. When the inter-stimulus interval is constant for a series of judgments, the probability of response increases as the interval decreases over the whole range of intervals used, as in previous experiments. But when the intervals are randomized over a 1-second range this increase is shown only as the longer intervals decrease, and at the shorter intervals a plateau appears instead. The 1-second range was chosen in the hope that it would produce an inflexion at about the middle of the range of inter-stimulus intervals; this appears to have occurred. In each condition P_M rises by 9·4 per cent between 2·75 and 1·75 seconds; from there to 0·25 seconds there is a rise of 9·0 per cent under the Constant Condition but of only 1·2 per cent in the Randomized Condition. Analysis of variance shows that the interaction between Interval and Condition is significant ($p < 0·05$). More positive responses were given in the Randomized than in the Constant Condition, but this variation is probably due to the procedure of choosing W and S stimuli at the beginning of each session on the basis of an initial estimation of the threshold.

TABLE II. PERCENTAGE OF POSITIVE RESPONSES FOR EACH CONDITION OF EXPERIMENT 2.
Inter-stimulus interval, or 'Sub-range' midpoint

(seconds):	0.25	0.75	1.25	1.75	2.25	2.75
<i>Constant Condition:</i>						
P_W :	50.0 ± 4.0	45.6 ± 3.9	35.0 ± 3.8	39.4 ± 3.9	35.0 ± 3.8	29.4 ± 3.6
P_S :	75.0 ± 3.4	78.8 ± 3.2	66.3 ± 3.7	67.5 ± 3.7	63.1 ± 3.8	58.8 ± 3.9
P_M :	62.5 ± 2.6	62.2 ± 2.6	50.7 ± 2.7	53.5 ± 2.7	49.1 ± 2.7	44.1 ± 2.7
<i>Randomized Condition:</i>						
P_W :	50.0 ± 4.0	51.3 ± 4.0	42.8 ± 3.9	48.9 ± 4.0	38.5 ± 3.9	40.0 ± 3.9
P_S :	73.1 ± 3.5	74.4 ± 3.5	78.1 ± 3.3	71.9 ± 3.6	65.0 ± 3.8	61.9 ± 3.8
P_M :	61.6 ± 2.6	62.9 ± 2.6	60.5 ± 2.6	60.4 ± 2.7	51.8 ± 2.7	51.0 ± 2.7
False Positives:						
Constant Condition:	13.1 per cent					
Randomized Condition:	5.2 per cent					

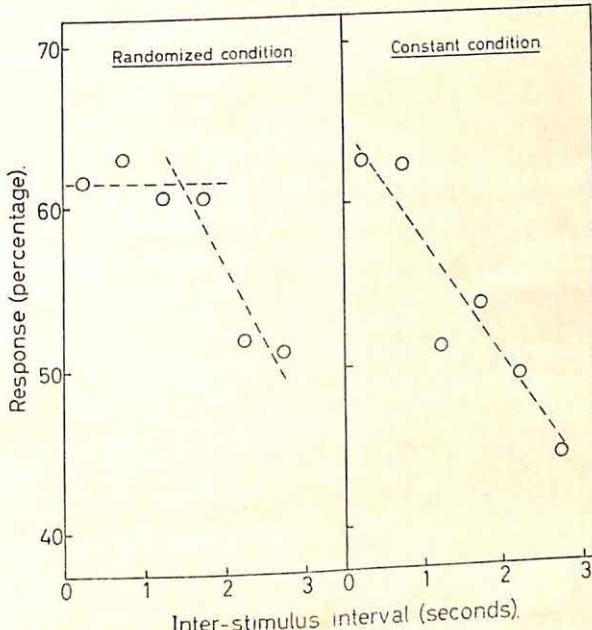


FIGURE 2. Mean frequencies of positive response expressed as percentages (P_M) to the W and S stimuli in Experiment 2 plotted against the inter-stimulus intervals in seconds. In the Constant Condition the length of the inter-stimulus interval was fixed for a series of trials. In the Randomized Condition it varied over a range of 1 second for each series; in the figure the result for each half-second 'sub-range' is plotted against its midpoint. The dashed lines (fitted by eye) indicate the general form of the results predicted, for the two conditions, by the range of expectation hypothesis.

V. PRELIMINARY THEORETICAL ANALYSIS

The results of the last two experiments, like those obtained by Howarth & Treisman (1958, 1961) and Treisman & Howarth (1959), appear to support the 'range of expectation' hypothesis rather than alternative explanations. If, then, we accept that the threshold is lowered for a period about the expected time of arrival of the stimulus, we must still consider whether this fall is inversely proportional to the duration of the range of expectation, as assumed by this hypothesis.

How long is a range of expectation? If Weber's law held exactly, and were applicable to the present problem, we might expect that for any inter-stimulus interval, T , the period of uncertainty about its end, during which the subject expects that the stimulus may be given at any moment and lowers his threshold accordingly, would be directly proportional to T . If this were so it would follow that the relation between threshold fall and range of expectation could be written as $(A-L) T = k$, where A is the asymptotic upper threshold level from which the threshold falls, L is the obtained threshold, T the inter-stimulus interval, and k a constant. It has been questioned whether Weber's law applies to time estimation (Woodrow, 1951); however, for intervals such as those under discussion, it has been shown to hold in the generalized form $\Delta T = c(T+a)$, where c and a are constants (Treisman, 1963 a). In this form Weber's law would lead us to predict

$$(A-L)(T+a)=k. \quad (1)$$

The law for temporal discrimination was determined in an experiment on time estimation in which the standard deviation of the productions or reproductions of an interval, T , was taken as the differential threshold, ΔT (Treisman, 1963 a). Subjects were required to produce or reproduce T by indicating the end of a time interval, rather than to detect the occurrence of a threshold stimulus, but otherwise the procedure was as similar as possible to that used by Howarth & Treisman (1958) in obtaining the thresholds shown in Figure 1. The law obtained was

$$\Delta T = 0.061(T+1.57). \quad (2)$$

These results make it plausible to assume that the range of expectation will be a linear function of inter-stimulus interval, but it does not necessarily follow that this function will include an additive constant with the magnitude of that in equation (2). The occurrence of the constant a in the linear generalization of Weber's law has sometimes been attributed to 'sensory noise' (Fechner, 1860; Gregory, 1956; Ekman, 1959). But 'noise' could arise not only in the input stage but also in comparison or judgment processes, or motor responses, required by the psychophysical tasks used to measure temporal discrimination. Thus the value of a might be less when the subject uses his 'internal clock' to determine the performance of a central process than when he is required to discriminate durations as such.

To test the applicability of equation (1) it was fitted to the five sets of data obtained by Howarth & Treisman (1958)—their Conditions 1, 2, 3, 4 and 7—which showed an effect of inter-stimulus interval on threshold, using the least squares criterion of best fit. The curves obtained for these data are shown in Figure 1. It can be seen that equation (1) gives a very good fit to the data, thus supporting the range of expectation hypothesis. It is also of interest that the additive constants are considerably smaller than those obtained in the time estimation experiments. For the five conditions, the values of a obtained were: -0.15 , -0.05 , -0.15 , 0.15 and 0.25 ; they are small and distributed around zero (their mean is 0.01). If in each case a is made equal to zero the resulting best-fitting curves are almost indistinguishable from those shown in Figure 1. If a is made equal to 1.57 sec., however, satisfactory fits are not obtained: in all five conditions the experimental points at the two extreme time intervals fall below these best-fitting curves, and most of the other points lie above them.

It seems we now have an answer to the first question posed in the introduction: the fall in threshold produced by an accessory stimulus can be attributed to the subject lowering his threshold during a range of expectation about the end of the inter-stimulus interval to an extent inversely proportional to the duration of the range. It also appears that this range is directly proportional to the length of the inter-stimulus interval, this relation not including the additive constant found in the Weber function resulting from temporal discrimination tasks. If we are to answer the second question we must now consider the mechanism of threshold change in more detail.

VI. SIGNAL DETECTABILITY THEORY

The most successful account of sensory discrimination has resulted from the application of statistical decision theory (Thurstone, 1927; Tanner & Swets, 1954; Gregory, 1956; Barlow 1957; Green, 1960; Swets, Tanner & Birdsall, 1961; Treisman, 1963 b), and it appears likely to be applicable here (Treisman & Howarth, 1959). Some elements of the theory relevant to the present problem will be briefly outlined.

Signal detectability theory treats sensory discrimination as a problem of distinguishing signals in noise. It is assumed that when a stimulus is presented to a subject it produces a central neural effect, x , which is used to determine the response. If a given stimulus is presented repeatedly the central effects produced, $x_1, x_2, \dots x_i, \dots x_n$, are described by a conditional probability density function $f_{SN}(x)$ which is normal with mean M_{SN} and standard deviation σ_{SN} ; this is the ‘signal + noise’ (SN) distribution. In the absence of the stimulus central effects occur which are described by a ‘noise’ (N) distribution, $f_N(x)$, with mean M_N and standard deviation σ_N . In the simplest case it is assumed that the two distributions have equal variance. If the subject must choose between two responses, ‘signal present’ (A), or ‘signal absent’ (CA), he will achieve optimal performance by comparing the likelihood ratio $\lambda(x_i)$ corresponding to the central

effect, x_i , produced on any trial (i), with a critical value of likelihood ratio, β , and reporting the signal present if $\lambda(x_i)$ is greater than β , and absent if it is less (Peterson & Birdsall, 1953; Birdsall, 1955). When likelihood ratio is monotonically related to x , then $\beta = \lambda(x_c)$, so that optimal decisions can be made by comparing obtained values of x directly with a critical value x_c . The probability of correct detections when the stimulus is presented, $P_{SN}(A)$, is then the integral of $f_{SN}(x)$ for $x > x_c$, and the probability of false positives in the absence of a stimulus, $P_N(A)$, is the corresponding integral for $f_N(x)$. (See Figure 3.)

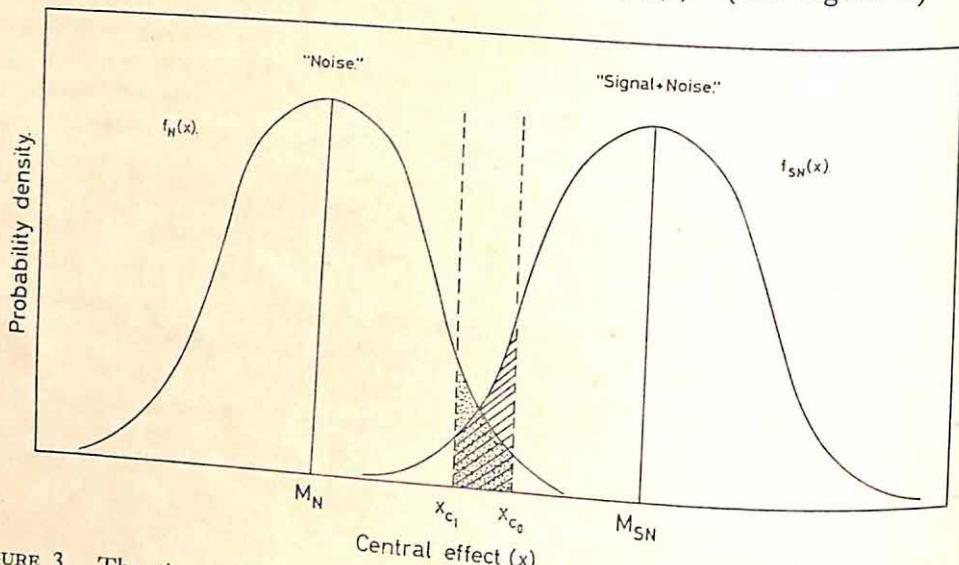


FIGURE 3. The signal detectability model: 'noise' and 'signal + noise' produce overlapping distributions of central effects.

The decision rule requires any central effect, x_i , to be compared with the critical value x_c . If $x_i > x_c$ a positive response is given, otherwise the response is negative. If x_{c_0} shifts to a new value x_{c_1} then $P_N(A)$, the probability of a positive response in the absence of a stimulus, will be increased by the dotted area, and $P_{SN}(A)$, the probability of detection when the stimulus is presented, will be increased by the area lined in.

The critical value of the likelihood ratio, β , and the corresponding critical value of the central effect, x_c , may be taken to be calculated by a 'criterion computer' (Swets, Tanner & Birdsall, 1955); how this is supposed to be done will depend on the definition of 'optimum performance' preferred by the theorist. There are a number of alternative ways of defining the optimal criterion, but the one that is usually preferred in applications of signal detectability theory is the criterion maximising the total expected value of the performance. This makes β a function of the *a priori* probabilities of signal and noise, $P(SN)$ and $P(N)$, and the values and costs attached to correct and incorrect positive and negative responses. However this definition is difficult to apply except to certain specially selected psychophysical procedures. To calculate β the subject must know *a priori* probabilities and values and costs;

to identify x_c he must know both $f_N(x)$ and $f_{SN}(x)$. These requirements restrict the model to procedures where the subject is well-informed; it cannot readily be applied to the discrimination of stimuli which are imperfectly known or unexpected. For example, when the method of limits is used the stimulus values presented depend partly on the subject's responses; he may not know the proper form of $f_{SN}(x)$ until the conclusion of the experiment. In this method, furthermore, $P(SN)=1$ and the maximal expected value criterion should then be $\beta=0$, but in practice very similar thresholds are obtained with this method and with procedures in which $P(SN) < 1$ (Treisman & Howarth, 1959). Consequently the theory is not usually applied to procedures such as the method of limits.

An alternative approach to the optimal criterion, that suggested by Neyman & Pearson (1933), has sometimes been used (Swets, Tanner & Birdsall, 1955). This requires $P_{SN}(A)$ to be maximised, subject to $P_N(A)$ not exceeding a constant limiting value, ϵ . β is then that value of $\lambda(x)$ for which

$$P_N(A) = \epsilon. \quad (3)$$

Since this definition requires less foreknowledge on the part of the subject it is more widely applicable. In order to determine x_c it is not essential for the subject to calculate β ; he need only be able to make some estimate of the value of $P_N(A)$ when he has adopted an arbitrary initial critical value of x (which he should ordinarily be well able to do, since daily life affords many opportunities to make difficult discriminations which can later be verified), and to vary this criterion until $P_N(A) = \epsilon$. The Neyman-Pearson definition is not in conflict with the finding that subjects may vary their criteria if monetary rewards or announced values of $P(SN)$ are altered (Swets, Tanner & Birdsall, 1961); such factors could act by altering the value of ϵ preferred by the subject. In the application of signal detectability theory to the present problems it will be assumed that the optimal criterion is computed in the Neyman-Pearson fashion.

VII. APPLICATION OF SIGNAL DETECTABILITY THEORY TO THE PRESENT PROBLEM

In applying a theory of this sort to the problem of analysing human performance three stages can be distinguished. In the first stage we must determine whether the theory adequately describes the overall relations between input and output. Signal detectability theory defines the course of action leading to optimal decisions when difficult discriminations must be made. We must, therefore, ask questions such as: Do subjects' decisions approximate to optimal decisions? When the experimental conditions vary does performance alter in the ways that would be expected if the theory properly describes the course of action being followed? Evidence such as the appropriate shape of subjects' ROC curves, etc., has tended to give positive answers to these questions, allowing us to accept the theory at this level. The second stage is the determination of the proper description of the operations the subject is actually performing: considered as a computer, he may not follow the signal detection programme in

precisely corresponding fashion; he may use approximate methods which although not strictly valid produce roughly equivalent input-output relations over the range of inputs encountered. At this level we know very little. In the third stage we attempt to identify at the physiological level the mechanisms whose modes of functioning have been described in more abstract form in the first and second stages. Here one identification can be made: it appears plausible that a physiological basis for 'noise', i.e. for the mean and variance of the 'noise' distribution, and the corresponding contributions to the SN distribution, may lie in the variable response and spontaneous activity of the sense organs and afferent pathways at the time of administration of the threshold stimulus (plus the physical variability of the stimulus) (Treisman, 1964). Sources of this sensory noise may include thermal agitation acting at receptors or afferent synaptic junctions (Buller, Nichols & Ström, 1953; Barlow, 1956); the arrival at afferent synapses of impulses originating outside the afferent system; and the spontaneous afferent discharges which have frequently been described (Granit, 1955).

A stimulus for which $M_{SN} = x_c$ would give $P_{SN}(A) = 0.5$, so that a 50 per cent threshold determined by the method of constant stimuli may be taken as an estimate of the stimulus value whose mean central effect corresponds to x_c . Similarly the threshold determined by the method of limits may be considered as a more or less accurate estimate of the stimulus corresponding to the criterion. Therefore the changes in threshold which were discussed above can be understood as consequences of changes in x_c . The signal detection model provides two ways in which such changes could be brought about: (1) The first, which will be referred to as a 'computation change' in criterion, would be caused by the output, β , of the 'criterion computer' changing to a new value, β' . This could result from a change in the information available to the computer, such as the *a priori* probabilities, values and costs, or the acceptable limiting false positive rate. β' would define a new critical value x_c' , given by $\beta' = \lambda(x_c')$. One feature of this sort of change is that if new information to the computer alters the criterion, the new value may be applied to determine the response to any central effect, x_i , which has occurred previously and been retained in a memory store from which it has not yet disappeared. Thus if an accessory stimulus acts by producing a computation change in criterion it could alter the frequency of response to a critical stimulus simultaneous with or even preceding it. (2) A 'distribution change' in criterion would result if, β remaining unchanged, one or both of the SN and N distributions were altered to give new functions, $f_{SN}'(x)$ and $f_N'(x)$. This would result in β defining a new critical value given by $\beta = \lambda(x_c')$. On the explanation of 'noise' given above a change in the means or variances of these distributions would be due to changes in the physiological variability or spontaneous activity of the afferent paths; x_c' would apply, then, to any value x_i' which could be taken to come from one of the new distributions $f_{SN}'(x)$ or $f_N'(x)$. Since a physiological change must take some time, however short, to be established, it follows that if the accessory stimulus acts by producing

distribution changes, then the fall in threshold which it causes must have some latency and could not occur when the accessory and critical stimuli are simultaneous; still less when the critical precedes the accessory stimulus.

A test of these alternatives is provided by experiments in which Treisman & Howarth (1959) examined the fall in threshold as the inter-stimulus interval was reduced from 1.5 to 0 seconds: the threshold fell continuously as the interval was decreased and was minimal when the stimuli were simultaneous. When the accessory stimulus followed the critical stimulus it continued to produce a significant fall in threshold up to intervals of 0.5-1.0 seconds. This shows that the subject's decision is not uniquely determined by his physiological state at the moment at which the stimulus is given; rather it appears that inputs can be stored for short periods, and if the accessory stimulus arrives before they have decayed the criterion can then be lowered and applied to them. It is of interest to note the similarity of the interval—0.5-1.0 seconds—for which inputs appeared to be stored in these experiments to that found for the decay of peripheral sensory information by Sperling (1960) in experiments on the span of apprehension.

The theory of signal detectability has now been applied to the accessory stimulus experiments, and it has been shown that in terms of this model the falls in threshold which were obtained can be attributed to computation changes in criterion rather than distribution changes. However, it still remains to see how a computation change in threshold could be related to the duration of the range of expectation in such a way as to produce the inverse relation between threshold fall and range of expectation found earlier, together with the constancy of false positive rate which was the second major problem noted in the introduction. Moreover it would further strengthen the case for applying the signal detectability model if it could be used to make quantitative predictions about the extent of threshold changes and these could be shown to agree with the data. These problems are pursued in the next section.

VIII. CHANGES IN CRITERION AND THE RANGE OF EXPECTATION

How could variation in the range of expectation affect the computation of x_c ? The signal detection model describes a decision procedure by which subjects can select answers to questions such as 'Was the signal present or absent?' It is assumed in applying the model that the subject makes one such decision for each response. But when there is a waiting time which affords more than one opportunity for the stimulus to occur then more than one decision might precede a response. Only if all the decisions were negative would the response be 'no'; if one or more decisions were positive the response would be 'yes'. In this case the false positive rate (*FPR*), i.e. the proportion of positive responses to 'blank' stimuli obtained experimentally, could not be taken as a simple estimate of $P_N(A)$, the probability that a value of x occurring centrally in the absence of a stimulus will exceed x_c , since on any trial several values of x may be tested, yet only one response is made. The question arises whether in this case the Neyman-Pearson criterion should be

taken to apply to $P_N(A)$ or to FPR , i.e. to all outputs of the signal detection mechanism, or to the overt decisions, the actual responses made in the experimental situation, only. The constancy of the false positive rate remarked on earlier indicates the latter, and this would have a pragmatic basis in subjects' need to avoid an embarrassing number of false perceptions during the course of their daily activities. Thus when the exact time of arrival of the stimulus is uncertain we will suppose that the subject's decision rule is: If any value of x obtained during the range of expectation exceeds x_c respond 'yes'; otherwise 'no'.

There is some evidence that subjects analyse their sensory input in successive 'sampling times' or 'moments' of about 100 msec. (Stroud, 1955; Cheatham & White, 1952, 1954; White & Cheatham, 1959); whether or not the input is segmented in this way, the 'moment hypothesis' facilitates exposition in the present problem, and to that extent will be accepted here. We will consider each waiting time to consist of a number of 'moments', at each of which a value of x is observed centrally and compared with the criterion; $P_N(A)$ is then the probability of a false positive arising from any comparison made at a moment when the stimulus was not presented. Since subjects must often, in the course of their daily activities, watch for stimuli whose appearance is uncertain, and whose time of arrival is indefinite, we may suppose that x_c is usually maintained sufficiently high to make the occurrence of an embarrassing number of false positives unlikely, even when watch must be maintained for some time. We may, therefore, ignore that part of the waiting time not contained in the range of expectation. If we suppose the latter to consist of r moments, for each of which $P_N(A) = \epsilon$, then the resulting false positive rate will be

$$FPR = 1 - (1 - \epsilon)^r. \quad (4)$$

We now have a basis for making predictions about the changes in the frequency of positive response to a stimulus (detection rate) that will take place as the inter-stimulus interval is varied. If the subject's efforts are directed to keeping FPR at an acceptable limiting rate then as the inter-stimulus interval, T , and consequently the number of moments, r , contained in the range of expectation vary, the subject will alter ϵ in a way intended to maintain FPR constant; this implies an inverse relation between ϵ and the length of the range of expectation. If at an inter-stimulus interval T_0 the range of expectation contains r_0 moments, for each of which $P_N(A) = \epsilon_0$, then if FPR is to have the same value at an interval T_1 , with range r_1 , ϵ_1 will be given by

$$\epsilon_1 = (1 - \epsilon_0)^{r_0/r_1}. \quad (5)$$

If the range of expectation is a constant proportion of the inter-stimulus interval, as was strongly suggested by the results discussed in Section V, then (except at very short intervals, when the range will approach the 'moment' in order of magnitude)

$$\text{Therefore } \epsilon_1 \text{ can be found provided } \epsilon_0 \text{ is known.} \quad r_0/r_1 = T_0/T_1. \quad (6)$$

If T_1 is shorter than T_0 , so that r_1 is less than r_0 , it follows that ϵ_1 will be greater than ϵ_0 , i.e. the criterion employed during the range of expectation will shift from x_{c_0} to a new lower value x_{c_1} . Figure 3 shows how the increase in ϵ producing this shift (represented by the dotted area of the 'noise' distribution) is accompanied by a corresponding increase in $P_{SN}(A)$ (the lined area of the SN distribution). It is apparent that if we know ϵ_0 and ϵ_1 , the corresponding shift in criterion can be calculated by taking the standardized normal deviates of the N distribution corresponding to these two probabilities, z_0 and z_1 , as measures of x_{c_0} and x_{c_1} . If we know $P_{SN}(A)_0$ we can then, on the assumption that the two distributions have equal variance, find the expected value of $P_{SN}(A)_1$ by consulting a table of the normal curve. For example, if $\epsilon_0 = 0.0227$ ($z_0 = 2.00$) and ϵ_1 is found to be 0.1587 ($z_1 = 1.00$), then if the detection rate for a stimulus is 50 per cent. when it is presented at T_0 , we would expect the rate to rise to 84 per cent. at T_1 . Since the stimulus is given at only one moment on each trial the detection rate can be taken as an estimate of $P_{SN}(A)$; this is not strictly accurate, since on trials on which the stimulus is presented some of the positive responses will nevertheless be due not to the central effect of the stimulus but to the effect of 'noise' at some other moment in the trial, so that a correction analogous to the 'guessing correction' should be made. But this effect is unlikely to be large and will be disregarded here.

Data suitable for comparing predicted and obtained detection rates are provided by the Constant Condition of Experiment 2, and by Experiments 1 (a) and 1 (b) of Treisman & Howarth (1959), since in each experiment the same stimuli were presented by the method of constant stimuli at more than one inter-stimulus interval, and most of these were not very short. Moreover the assumption that the 'noise' and 'signal + noise' variances are equal is plausible for these data since σ_N represents the variability of the central response to a tone 60–70 db. above threshold, and σ_{SN} that of the response to a tone a fraction of a decibel louder. In order to derive predictions the value of ϵ must be known for at least one inter-stimulus interval. A value was estimated, somewhat arbitrarily, by taking $T_0 = 10$ seconds and supposing that any change in ϵ during the course of this interval was negligible. Taking FPR as 4.1 per cent, the average value in the experiments to be examined, and making $r = 100$ (the number of moments in T_0 , using Stroud's (1955) estimate of 100 msec. for the moment) in equation (4), ϵ_0 was found to be 0.00043. Using equations (5) and (6), values of ϵ and z for other time intervals were computed. For each stimulus strength the frequency of positive response at the longest interval in the experiment was taken as the starting-point, and the detection rates at other intervals were predicted from the shifts in z . These predictions are shown as continuous curves in Figure 4, for the Constant Condition of Experiment 2, and in Figure 5 for stimuli used in Experiment 1 of Treisman & Howarth (1959). Each response percentage is based on 160 responses in Experiment 2, and 240 for the 1 and 8 second intervals, and 80 for the 3 and 6 second intervals in Figure 5. Allowing for error variation, the agreement between

these predictions and the experimental results appears satisfactory. Twelve experimental points are greater than their predicted values and 9 are less, a difference which is not significant ($\chi^2 = 0.43$). The assumptions made in deriving ϵ_0 were not critical for producing this agreement: varying ϵ_0 between 0.0001–0.001 did not markedly impair it. In making these predictions it was assumed that $a=0$, as found in Section V. (Predictions were also made for $a=1.57$ sec. but were clearly unsatisfactory: the rises in detection rate were underestimated for 19 of 21 experimental points.)

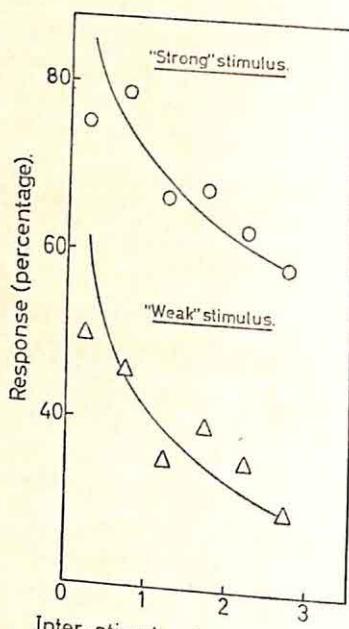


FIGURE 4. Mean detection rates for the *W* and *S* stimuli in the Constant Condition of Experiment 2. The continuous curves show the theoretical detection rates predicted in Section VIII.

We have now combined the signal detection model with the range of expectation hypothesis and shown that, taken together, they allow the magnitude of the changes in detection rate with change in inter-stimulus interval to be predicted with fair accuracy, on the basis that subjects tend to keep their false positive rates constant. We must still relate this result to the inverse relation between threshold fall and range of expectation found in Section V. It was seen that values of the threshold obtained experimentally can be considered to estimate the stimulus values corresponding to different levels of x_c . In Figure 6 the values of ' z_1 ' (estimates of x_c expressed as standardised normal deviates of the *N* distribution) used in making the predictions in Figures 4 and 5 have been plotted for the inter-stimulus intervals between 1 and 9 sec. It can be seen that their relation to the inter-stimulus intervals is similar to that shown by the threshold data plotted in Figure 1. The dotted line shows the

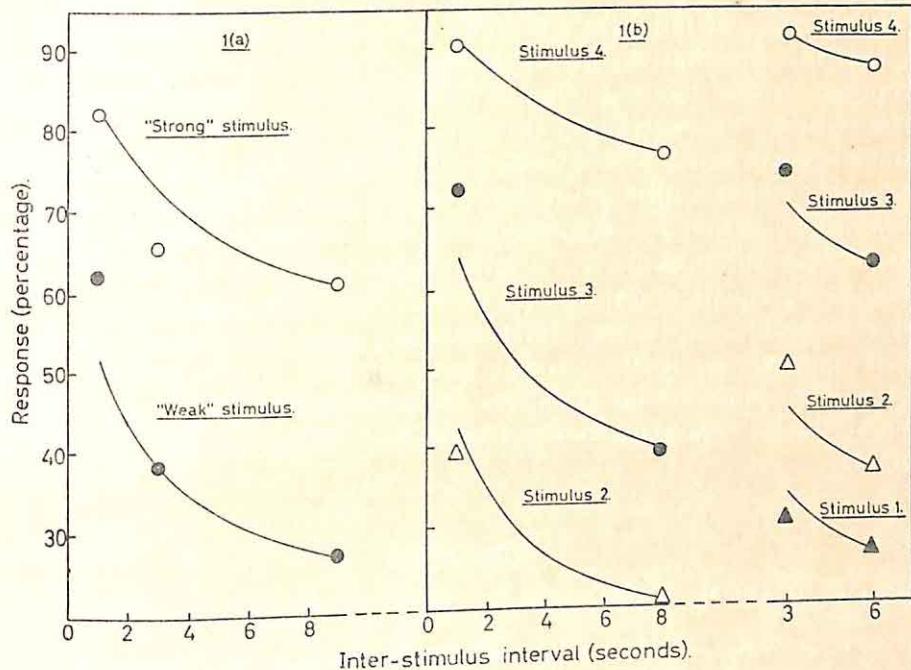


FIGURE 5. Data from Treisman & Howarth (1959) Experiment 1. Detection rates (frequencies of positive response when the stimulus is presented, expressed as percentages) are shown for stimuli of constant magnitude at different inter-stimulus intervals. The continuous curves show the rates predicted in Section VIII.

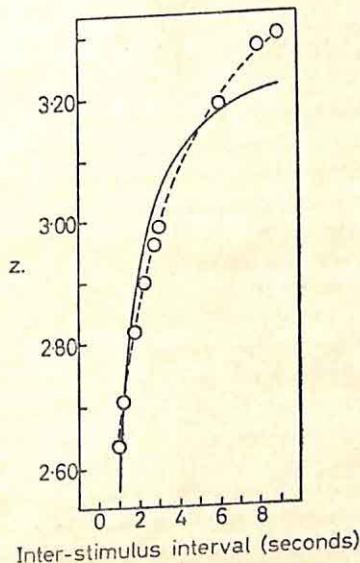


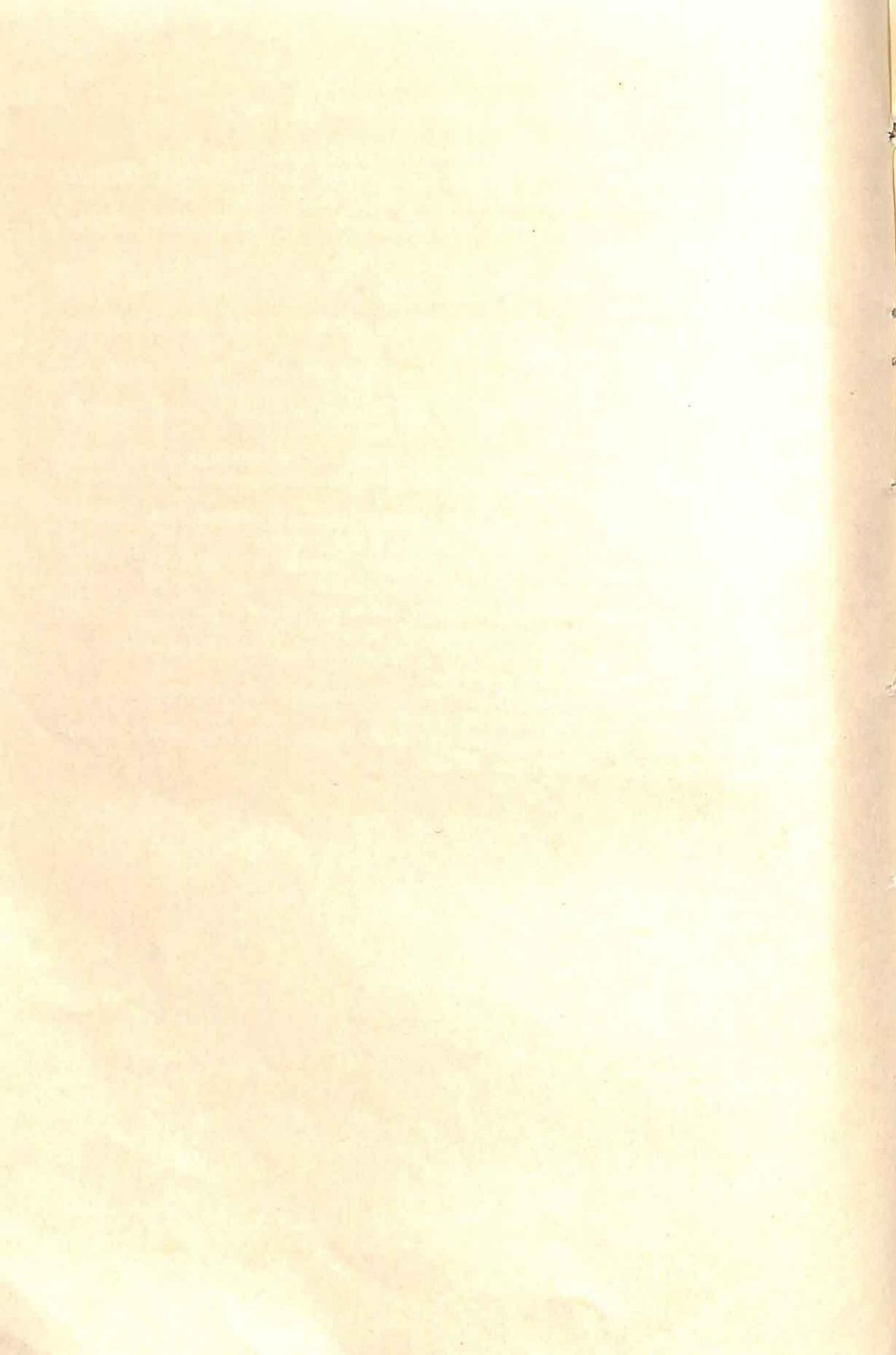
FIGURE 6. The values of z used in making the predictions shown in Figures 4 and 5 are plotted against the corresponding inter-stimulus intervals for the range 1–9 seconds. The dashed line is the best-fitting curve of the form $(A-z)(T+a)=k$; it is given by $3.75-z=3.12/(T+2.36)$. The continuous line is the best-fitting function when $a=0$.

best-fitting curve of the form $(A-x)(T+a)=k$ (compare equation (1)). It is clear that this function would supply very good estimates of the values of x_c appropriate to the different inter-stimulus intervals within this range. It is less curved than the best-fitting curves in Figure 1, since it has $a=2.36$. The continuous curve is the best fit when $a=0$. It would give good but not optimal values of x_c , and is similar to the curves in Figure 1. This brings us to a suggestion which is more an hypothesis for future research than a conclusion: when the inter-stimulus interval is varied the subject aims at optimal behaviour, as defined by equations (4) and (5) and the associated calculations; however he avoids performing these rather complex computations by employing a simpler approximate method which gives roughly similar answers for this range of inter-stimulus intervals: he computes critical values of x from the relation $(A-x_c) T=k$, giving A and k values which maintain FPR approximately constant.

REFERENCES

- BARLOW, H. B. (1956). Retinal noise and absolute threshold. *J. opt. soc. Amer.*, **46**, 634-639.
- BARLOW, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *J. Physiol.*, **136**, 469-488.
- BIRDSELL, T. G. (1955). The Theory of Signal Detectability. In: H. Quastler (ed) *Information Theory in Psychology*. Glencoe, Ill.: The Free Press.
- BULLER, A. J., NICHOLLS, J. G., & STRÖM, G. (1953). Spontaneous fluctuations of excitability in the muscle spindle of the frog. *J. Physiol.*, **122**, 409-418.
- CHEATHAM, P. G. & WHITE, C. T. (1952). Temporal numerosity: I. Perceived number as a function of flash number and rate. *J. exp. Psychol.*, **44**, 447-451.
- CHEATHAM, P. G. & WHITE, C. T. (1954). Temporal numerosity: III. Auditory perception of number. *J. exp. Psychol.*, **47**, 425-428.
- EKMAN, G. (1959). Weber's law and related functions. *J. Psychol.*, **47**, 343-352.
- FECHNER, G. T. (1860). *Elemente der Psychophysik*. Breitkopf & Hartel: Leipzig.
- FINNEY, D. J. (1952). *Probit Analysis*, 2nd Ed. Cambridge: Cambridge University Press.
- FISHER, R. A. (1954). *Statistical Methods for Research Workers*. London: Oliver & Boyd.
- GRANIT, R. (1955). *Receptors and Sensory Perception*. New Haven: Yale University Press.
- GREEN, D. M. (1960). Psychoacoustics and detection theory. *J. acoust. soc. Amer.*, **32**, 1189-1203.
- GREGORY, R. L. (1956). An experimental treatment of vision as an information source and noisy channel. In: E. C. Cherry (ed). *Information theory. Third London Symposium*. London: Butterworths Scientific Publications.
- HOWARTH, C. I. & TREISMAN, M. (1958). The effect of warning interval on the electric phosphene and auditory thresholds. *Quart. J. exp. Psychol.*, **10**, 130-141.
- HOWARTH, C. I. & TREISMAN, M. (1961). Lowering of an auditory threshold by a near threshold warning signal. *Quart. J. exp. Psychol.*, **13**, 12-18.
- HUMPHREY, G. (1933). *The Nature of Learning*. London: Kegan Paul, Trench, Trubner & Co. Ltd.
- NEYMAN, J. & PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. A.*, **231**, 289-337.
- OLDFIELD, R. C. (1937). Some recent experiments bearing on 'Internal Inhibition'. *Brit. J. Psychol.*, **28**, 28-42.

- PAVLOV, I. P. (1927). *Conditioned Reflexes*. Oxford: Oxford University Press.
- PETERSON, W. W. & BIRDSALL, T. G. (1953). The theory of signal detectability. Technical Rep. 13, Electronic Defense Group, University of Michigan.
- SPERLING, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.*, 74, Whole No. 498.
- SWETS, J. A., TANNER, W. P. & BIRDSALL, T. G. (1955). The evidence for a decision-making theory of visual detection. Technical Rep. 40, Electronic Defense Group, University of Michigan.
- SWETS, J. A., TANNER, W. P. & BIRDSALL, T. G. (1961). Decision processes in perception. *Psychol. Rev.*, 68, 301-340.
- STROUD, J. M. (1955). The fine structure of psychological time. In H. Quastler (ed) *Information Theory in Psychology*. Glencoe, Ill.: The Free Press.
- TANNER, W. P. & SWETS, J. A. (1954). A decision-making theory of visual detection. *Psychol. Rev.*, 61, 401-409.
- TANNER, W. P., SWETS, J. A. & GREEN, D. M. (1956). Some general properties of the hearing mechanism. Technical Rep. 30, Electronic Defense Group, University of Michigan.
- THURSTONE, L. L. (1927). Psychophysical analysis. *Amer. J. Psychol.*, 38, 368-389.
- TREISMAN, M. & HOWARTH, C. I. (1959). Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quart. J. exp. Psychol.*, 11, 129-142.
- TREISMAN, M. (1963 a). Temporal Discrimination and the indifference interval: implications for a model of the 'internal clock'. *Psychol. Monogr.*, 77, Whole No. 576.
- TREISMAN, M. (1963 b). Auditory unmasking. *J. acoust. soc. Amer.*, 35, 1256-63.
- TREISMAN, M. (1964). Noise and Weber's Law: The discrimination of brightness and other dimensions. *Psychol. Rev.* (in press).
- WHITE, C. T. & CHEATHAM, P. G. (1959). Temporal numerosity: IV. A comparison of the major senses. *J. exp. Psychol.*, 58, 441-444.
- WOODROW, H. (1951). Time Perception. In: S. S. Stevens (ed): *Handbook of Experimental Psychology*. New York: Wiley.



A METHOD OF TREATING INDIVIDUAL DIFFERENCES IN MULTI-DIMENSIONAL SCALING*

By J. A. KEATS
University of Queensland

Three studies are reported in which results obtained by analysis using the multi-dimensional 'unfolding' method are compared with those obtained by multi-dimensional scaling. Attitudes of undergraduates towards crime, political parties and types of accommodation were assessed by presenting the stimuli two at a time and asking subjects to indicate preference as well as rate the differences between each pair of stimuli. Subjects were also asked to indicate whether they liked, disliked or felt indifferent towards each of the stimuli. Samples of approximately 200, 500 and 1,000 subjects respectively were used in the three studies. Where the individual differences in attitudes were relatively small, as in the crimes data, the two methods gave equally adequate representations of the data. However in their attitudes towards political parties the students showed large individual differences and this was reflected in considerable disparity between the results obtained by the two methods. An attempt was made to combine the two methods to investigate distortion produced by strong negative attitudes and the results obtained were linked to a theory of psychological balance. This theoretical explanation seemed appropriate to the results obtained from the study of students reactions to different types of accommodation.

I. HISTORICAL BACKGROUND

The method of paired presentations envisages a situation in which two stimuli are presented to a subject and he is asked to indicate which is heavier, more desirable, preferred by him, etc. In certain situations the data obtained can be regarded as proximity data. For example, if the subject says he prefers stimulus A to stimulus B, then he can be regarded as saying that in a psychological sense, stimulus A is closer to him than stimulus B. One of Thurstone's first illustrations (1927) of analysis by the method of paired comparisons required male and female students to indicate which of a pair of crimes was more serious in his or her view. If each student can be thought of as reflecting the group standards to the same extent, then the resulting scale values indicate the degree of rejection of the various crimes by the group as a whole.

Thurstone noted that there was an inconsistency in his data which could not be accounted for by his method of analysis. When the two crimes, rape and homicide were presented, homicide was judged to be the more serious by the majority (55.9%) of the cases. On the other hand, when scale values were computed, rape had a higher value (3.275), i.e., was regarded as more serious than homicide (3.156). He left this paradox unresolved in his original article.

* Based on a paper presented at the International Congress of Psychology, Washington, 1963.
This research was supported in part by the Social Science Research Council of Australia.

Some of Thurstone's data are reproduced in Table I. If the percentage who judge rape to be more serious than a particular crime is greater than the percentage judging homicide to be more serious than the same crime, then rape will have the higher scale value in that pair of comparisons. This is the situation to a marked degree in the case of the three sex crimes, abortion, adultery and seduction. The fact that these three give the largest difference in favour of 'rape' suggests that sex crimes of which rape is one are being judged on a different dimension from the one operating with the other crimes. Thus the inconsistency observed is explicable in terms of a two dimensional space rather than the single dimension usually assumed in applications of the method of paired comparisons. There is clearly a need for multidimensional analysis of these proximity data.

TABLE I. PERCENTAGE OF OCCASIONS ON WHICH HOMICIDE AND RAPE ARE JUDGED TO BE MORE SERIOUS THAN OTHER CRIMES (Thurstone, 1927)

	<i>Homicide</i>	<i>Rape</i>
Abortion	76.0	81.2
Adultery	86.3	92.5
Seduction	81.9	92.4
Rape	55.9	(50)
Kidnapping	91.7	90.2
Assault	97.0	94.7
Burglary	98.1	98.1
Vagrancy	98.9	98.5

A further assumption of the method of paired comparisons is that each subject is attempting to make the same judgement and that differences which are observed are due solely to fluctuations in the discriminative process. In such cases it is most unlikely that a particular subject's responses will generate a consistent ordering of the stimuli but rather that inconsistencies will occur because of these fluctuations in the subject over time. In an actual experiment in which stimuli were presented two at a time it would be a simple matter to test each subject's responses for consistency. If it were found that a high percentage of subjects was consistent but that groups of subjects gave radically different consistent orderings, this could be taken as evidence of individual differences and analysis by Thurstone's method of paired comparisons would be inappropriate. In Thurstone's example one might expect individual differences in attitudes towards particular crimes—for example, women students are likely to consider sex crimes more serious than do men students. For this reason a useful further extension of the method of paired comparisons would be one which would examine the data for consistency of each subject's response as an indication of the presence of systematic individual differences. If such differences were established in a particular case, they should be taken into account in the scaling method.

Various unidimensional scaling methods were developed and used during the 1930's and Guttman developed his procedure for scaling persons and items in the one operation, i.e., took account of individual differences in scaling items. The basis of extending scaling to the problem of establishing more than one dimension was also developed in that period although not applied extensively until after Torgerson (1952). Towards the end of the 1940's Coombs (1952) developed a method of representing persons and stimuli in the same single dimension using ordinal methods. This method was extended to more than one dimension by Bennett and Hays (1960). Tucker and Messick (1963) have recently extended the method of Torgerson to the case in which individual differences can be taken into account and their effect on scale values assessed. Sheppard (1962) has presented an alternative method of analysing ordered distances. The purpose of the present paper is to illustrate the sort of result obtained by the multidimensional analysis of rank orders following the method of Bennett and Hays, and to compare this with that obtained by the alternative multidimensional analysis of Torgerson.

II. PREFERENCE AND JUDGED DISTANCE

It must be conceded that logically there is no necessary connection between the Torgerson method and the multidimensional unfolding method of Bennett and Hays since the rank order analysis depends upon the existence of relatively large individual differences and ideally the quantitative analysis requires that individual differences be as small as possible. The former depends on the notion that persons and stimuli can be arranged so that the order of the distances from each person to the stimuli corresponds to the ordering of the stimuli by the person. The quantitative method depends upon the notion that the person can judge the distance between each pair of stimuli, or more simply, can decide whether the distance between stimuli A and B is greater or less than the distance between stimuli C and D irrespective of the location or identification of the person. Even though there is no logical necessity for results using these two methods to be related, it is still a matter of empirical interest to discover if they are, in particular cases. It is in this spirit that the following studies are presented.

In the first study, six of the nine stimuli of Table I were selected and the 15 possible pairing were presented to each of 250 subjects. The stimuli used were rape, murder (homicide), adultery, seduction, assault and burglary. Burglary and assault were combined and presented as one stimulus to be compared with pairings of other stimuli, some of which will be mentioned below. This was done to try to force the subject to concentrate on the particular pair and not to relate this decision to any he had previously made. In this way it was hoped to ensure that any consistency in the orderings was a function of the person and not simply an expression of his desire to appear consistent. In each case the subject was asked to indicate which of the two stimuli he preferred in the sense of liked best or disliked least. In addition the subject was asked to rate the distance

between the stimuli on factors which affected his preference. The data were analysed in three ways:

1. Thurstone's method of paired comparisons (case V) whereby each proportion of occasions on which stimulus A is chosen over stimulus B is transformed by the cumulative normal distribution and averaged for each stimulus to give a scale value (see Torgerson, 1958, p. 159ff.).

2. The ratings of distance were converted to scale values using the method of successive intervals (see Torgerson, 1958, p. 227ff.). After the addition of the appropriate constant to make them positive, these distances values were converted to scalar products and factored following Torgerson (1952 and 1958, p. 280ff.).

3. Where subjects gave completely transitive preferences, these could be converted to a unique ordering (i.e. with no ties). From these orders decisions as to the dimensionality of the stimuli and their location within these dimensions could be made following Bennett and Hays (1960) and McElwain and Keats (1961).

The results obtained from the first analysis are given in Table II. It will be noted that the women thought sex crimes more serious than did the men. The Queensland data are in other ways comparable to Thurstone's but the discrepancy noted by Thurstone does not occur for rape and murder.

TABLE II. SCALE VALUES OBTAINED BY THURSTONE COMPARED WITH THOSE OBTAINED ON A QUEENSLAND GROUP

	Thurstone	Queensland Men	Queensland Women
Burglary	1.51	1.26	0.87
Assault	1.47	2.06	2.00
Adultery	2.10	1.90	2.24
Seduction	2.29	1.41	2.06
Rape	3.28	3.25	3.25
Homicide (Murder)	3.16	3.91	3.36

The multidimensional analysis of distances yielded three dimensions in the case of the six stimuli, but when burglary and assault were combined only two dimensions were necessary. This example is presented because of

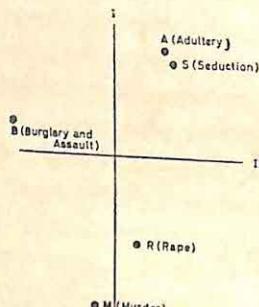


FIGURE 1. Multidimensional Scaling of Five Crimes. (See text for detail.)

its relative simplicity—see Figure 1. It may be noted that the second factor distinguishes the sex crimes from the remainder.

Finally the data were converted to rank orders and it was found that approximately 80 per cent of cases gave consistent orders. Under the randomization hypothesis with six stimuli the probability of a consistent ordering is $6!/2^{15} = 0.22$ or 2.2 per cent and with five stimuli this percentage rises to 12 (approximately). With the number of cases involved (250) it is obvious that 80 per cent is significantly greater than 12 per cent.

The basis of the multidimensional unfolding method for rank orders is the same as that for the unidimensional analysis i.e. that stimuli are arranged so that the orders that do occur in subjects' responses can occur in the diagram. For example with three stimuli A, B and C if the two possible orders with B ranked last are never given by subjects, then the stimuli might be represented on a straight line with B in the middle. From Figure 2 (a) it can be seen that with such an arrangement points to the left of the midpoint of A B correspond to the order A B C since the points A B and C are in that order of proximity from all points to the left of the mid-point of A B. Similarly the orders B A C, B C A and C B A can occur in this figure but *not* the orders A C B and C A B. However if the point B is lowered so that now the three points define a plane and the perpendicular bisectors of A B and B C are extended to meet as in Figure 2 a, then these two orders, A C B and C A B do occur. Thus with three stimuli in two dimensions all of the six possible orders are available. With four stimuli in two dimensions there is again some implied restriction on the orders given by subjects and a corresponding restriction on the arrangement of the stimuli. For example, if one point lies within the triangle of the other three it cannot be placed last (McElwain and Keats, 1961). The order corresponding to a particular region is the order of the proximities of the stimuli to each point in the region.

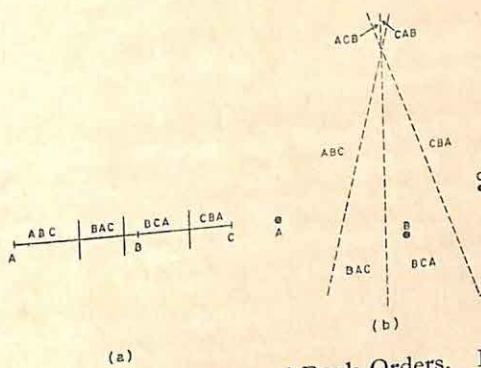


FIGURE 2. Multidimensional Representation of Rank Orders. In (a) objects A, B and C can be placed on a single dimension. The segments of the lines bounded by the vertical lines indicate points (i.e. persons) which have distances from the three objects which are in the indicated order. In (b) B has to be placed below A and C, and perpendicular bisectors of AB and BC are drawn to permit the representation of the orders ACB and CAB which cannot be achieved in (a).

Both the six stimuli and the five stimuli orderings of the crimes could be fitted fairly adequately into two dimensions using this principle, with over 90 per cent of the consistent orderings accounted for in the latter case. There was perhaps a suggestion of a third dimension, but many more cases would be required before this could be regarded as established. Figure 3 indicates a location of stimuli and persons which would account for as many orderings as possible. The lines are the perpendicular bisectors of the lines joining the points, and each area enclosed by these lines corresponds to a particular order of distances from a point in this area to the stimuli, i.e. to a particular ordering given by a subject.

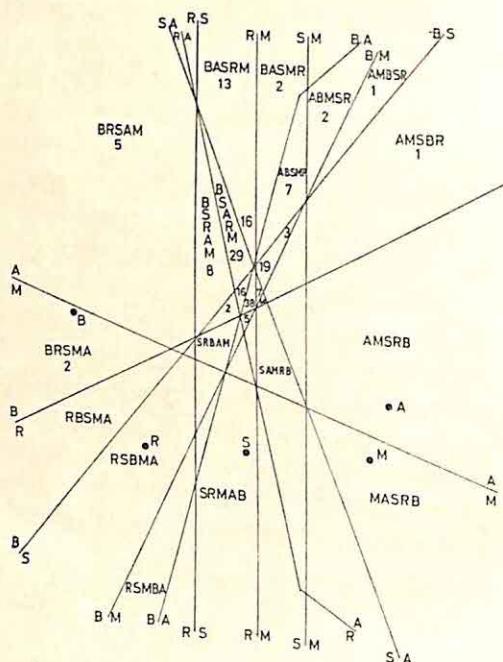


FIGURE 3. Multidimensional Analysis of the Ranking of Five Crimes. The figures in each area indicate the number of subjects exhibiting the corresponding order.

By drawing in the ten perpendicular bisectors of the lines joining pairs of points in Figure 1, it is possible to discover the rank orders generated by this figure. These orders account for 80 per cent of the subjects' orders. Thus in terms of orderings, the Torgerson solution accounts for the data almost as well as the best two dimensional unfolding solution. The subjects perceive the distances between stimuli in such a way that they will generate the orderings of almost the whole group. From the point of view of the subjects' orders of preference, there has been little distortion in the perceived arrangement.

III. CATEGORIZING SUBJECTS USING RANK ORDER DATA

The analysis of rank orders provides a basis for dividing the original group of subjects into sub-groups. However, although neighbouring subdivisions

correspond to different orderings they may not correspond to marked differences in feeling towards stimuli. To examine this point more carefully the subjects were also asked to indicate whether they liked, disliked or felt indifferent towards the objects or activities represented by the stimuli. Neighbouring groups could then be compared to determine whether their slight difference in ordering corresponds to a large or small difference in liking. If this difference is large in some sense then the boundary may be thought of as strong since it separates people with quite different views. In the case of the crimes only a very small section of a few of the boundaries could be thought of as strong in this sense. However in a study reported elsewhere (Keats, 1962) the stimuli 'saints', 'artists', 'heroes' and 'scholars' were used and the percentages liking each of the stimuli for each of the determined rank orders were calculated. Neighbouring regions could thus be compared by assessing the significance of difference between these percentages. If this difference is significant this fact can be taken as evidence that a strong boundary separates the two regions. In this case there was a strong boundary, separating those who put 'saints' last from the rest. When the rated distances were analysed multidimensionally for the contrasting groups, i.e. those who put 'saints' first and those who put them last, quite different two-dimensional arrangements were obtained. The points for the "saints first" group accounted for a high percentage of the orders of the whole group when interpreted in terms of the multidimensional unfolding model, but that of the "saints last" group accounted for only 60 per cent of the total orders. What was not noted at that time was that this 60 per cent was made up of 100 per cent of the orders for people who put 'saints' last but only 49 per cent of the orders for people who did not. This suggests a possible generalization, i.e. that people who are on the negative side of strong boundaries tend to distort their perception of the distances between stimuli in such a way as to preserve their own orderings but to eliminate orderings conflicting with these. This generalization will be examined with respect to a third set of stimuli.

The third set of stimuli to be considered was used in the same study and also with a subsequent group so that approximately 500 subjects were available. The stimuli consisted of the five political parties regularly contributing candidates for the Queensland State elections, i.e. Liberal Party (L), Country Party (C), Australian Labor Party (A), Queensland Labor Party (Q) and Communist Party (K). The analysis of rank orders gave a very good fit with two dimensions as shown in Figure 4, but there were many strong boundaries indicating that even a minor change in the order of political parties can be associated with considerable feeling. This of course was particularly true when the first choice was involved. In this example it was possible to isolate a quite homogeneous group of Liberal voters and a fairly homogeneous group of Labor voters. The distance ratings were analysed separately for these two groups and the results are presented in Figure 5. These ratings indicated two dimensions when the two groups were treated separately, but the ratings for the whole group required four dimensions. There is a basic similarity in these two representations

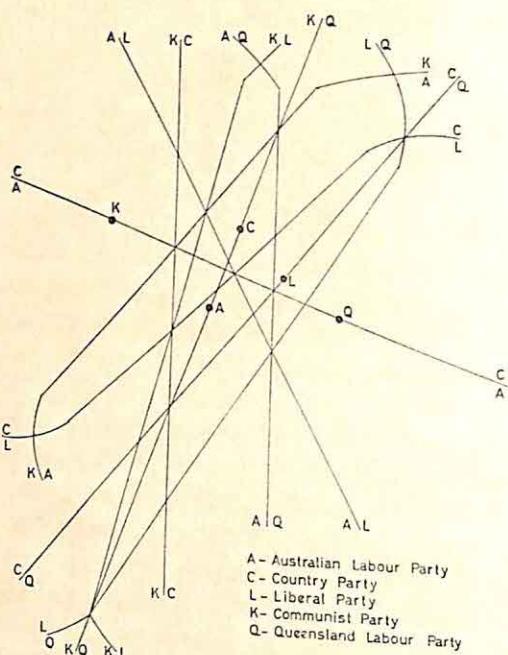


FIGURE 4. Multidimensional Analysis of Five Political Parties. Lines which intersect off the page have been curved to meet on the page, but these curves must not be produced to intersect with other lines; for example the perpendicular bisector of AK does not meet that of CQ as these lines are parallel.

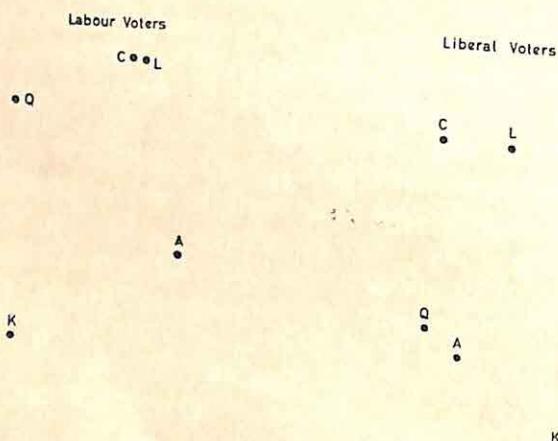


FIGURE 5. Multidimensional Scaling of Five Political Parties.

with the Country Party shown as furthest from the Communist Party whereas the analysis of rank orders shows the Queensland Labor Party as in greatest contrast to the Communist Party. Ideologically the latter would appear to be the more correct representation as the Queensland Labor Party is violently anti-Communist in its policy. The difference between the two diagrams arises mainly in the location of this party. The Liberal voters see it as belonging to what might be called the socialist group but the Labor voters see it as grouped with the Liberal and Country parties. This distortion is more understandable if it is realized that the Queensland Labor Party had been in existence for only a few years when this survey was made. It is rather interesting to note that both of the major parties disown it.

This example is an excellent one for testing the generalization suggested earlier, i.e. that subjects on the negative side of strong boundaries tend to distort their perceptions of the distances between stimuli in such a way as to preserve their own orderings but to eliminate orderings conflicting with these. The two representations given in Figure 5 may be interpreted in terms of the unfolding model by drawing the 10 possible perpendicular bisectors of the lines joining the 10 possible pairs of points. The bisectors will generate regions corresponding to one of the possible ways of ranking the five points in their orders of proximity to any point in the region. In this way a maximum of 46 orders can be generated by each of the arrangements but since the arrangements differ in their location of the Queensland Labour Party and in other ways they will generate some different rank orders. These two sets of orders may then be compared with those given by the subjects to discover how closely they correspond. It would be expected from the generalization that the arrangement for the Liberal voters would exclude some of the orders (and thus some of the subjects) which have the Australian Labor Party in first position and vice versa. The location of stimuli by the Liberal Party voters accounted for 83 per cent of the orders and 99 per cent of the cases who put the Liberal Party first but for those who put the Labor Party first only 33 per cent of the orders and 29 per cent of cases were accounted for. The corresponding figures for Labor voters were 78 per cent of the orders and 85 per cent of the cases who put the Australian Labor Party first but only 25 per cent of the orders of people who put the Liberal Party first even though these accounted for 87 per cent of such cases. The generalization seems to hold quite well for this example when orders are considered but breaks down somewhat for individual cases because of the high frequencies in two of the 12 orders with the Liberal Party first.

A further example illustrates one way in which the diagrams obtained by the analysis of rank orders can be influenced by personal identifications. In a survey of accommodation needs, students at the University of Queensland were asked to rank the four possibilities, living at home, in a University College, in a flat or apartment, and in a boarding house in order of their preference. There were approximately one thousand students interviewed and an initial inspection of the orders given revealed that all twenty-four possible orders

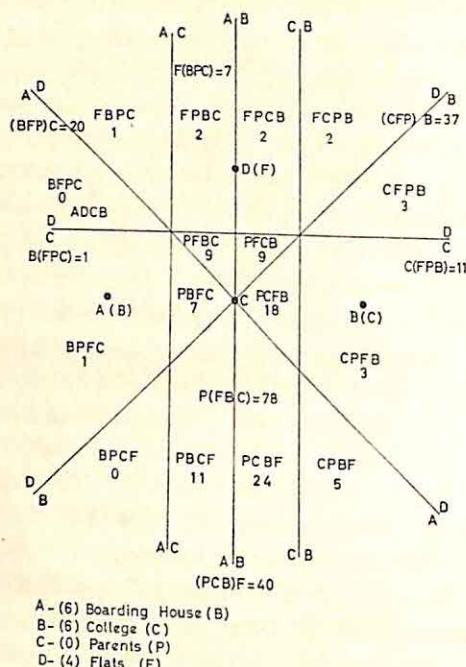


FIGURE 6. Accommodation Preferences of Students living at Home. The number of subjects exhibiting each order is shown. Also the number placing an alternative P. first or last is given. E.g., the equation $P(FBC)=78$ gives the number favouring P.

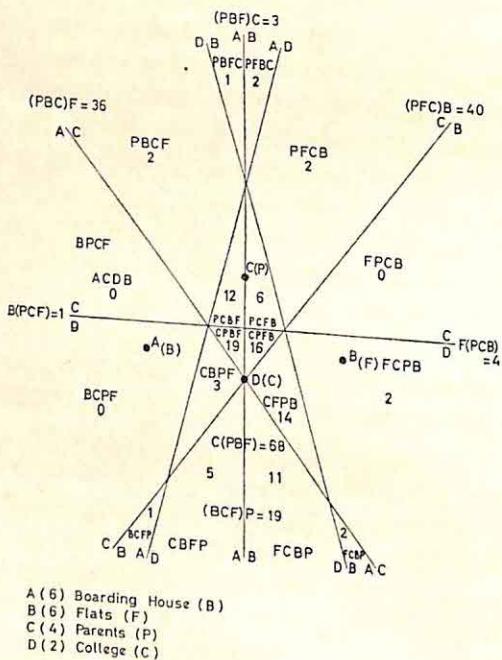


FIGURE 7. Accommodation Preferences of Students living in Colleges.

Treating Individual Differences in Multi-dimensional Scaling

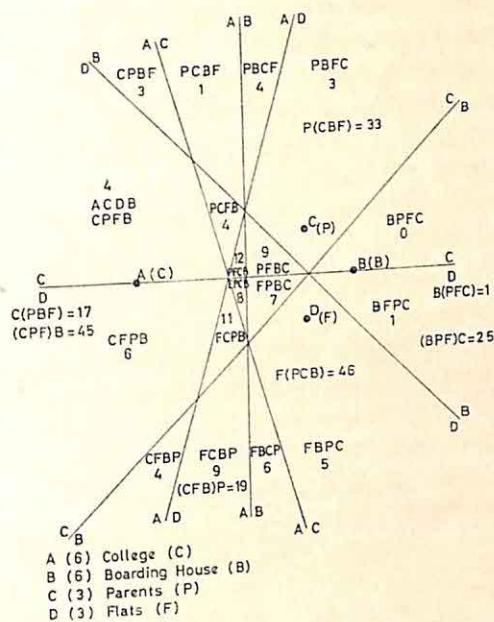


FIGURE 8. Accommodation Preferences of Students living in their own House or Flat.

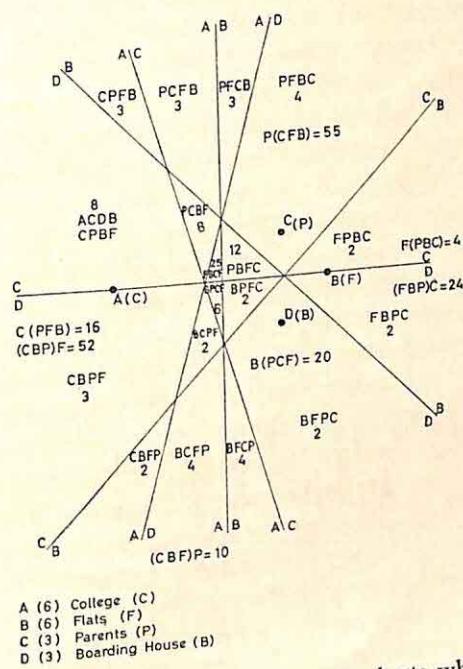


FIGURE 9. Accommodation Preference of Students who are Boarding.

were well represented, i.e. that three dimensions would be required to represent the entire data. When the data were classified according to where the student actually lived it was found that representation in two dimensions accounted for the orderings of each group very well, but a different diagram was needed for each one. In fact there was a highly significant Chi square for a comparison of the frequencies in the critical areas distinguishing the various diagrams. An examination of Figures 6-9 reveals that for students living at home, this is the central stimulus, but for students living in other types of accommodation, their particular type is contrasted with home. There were no notable age or sex differences in type of diagram, and in fact the Asian men present in the sample showed choice patterns which conformed to the diagrams presented, just as well as the Australians. Thus two of the three dimensions necessary to describe these data appear to be the actual place of living and the contrast between this and home. The third dimension contrasts University College with flats and boarding houses. Closer examination of these diagrams reveals very clearly that University Colleges provide the most satisfactory alternative form of accommodation for students unable to live at home.

IV. APPLICATION TO HEIDER'S THEORY OF BALANCED RELATIONSHIPS

The primary purpose of the present studies was to obtain data which would enable the results obtained by different scaling procedures to be compared. These results have intrinsic interest, but in this case they also contribute to what is probably the primary function of scaling theory, i.e. to provide quantitative procedures by which psychological theories can be assessed and extended. Berger, Cohen, Snell and Zelditch (1962) have formalized Heider's theory of balanced relationships between a given person and other persons or objects. This formalization related persons and objects by dynamic relationships, L, e.g. like, respect, etc. and unit relationships, U, e.g. owns meets, works with, etc. and their opposites $\sim L$ and $\sim U$. In general a system of persons and objects is balanced for a person, for example if all his relationships are positive and if he sees all other persons' relationships as positive. Berger *et al.* (1962, pp. 14-15) quote four propositions from Heider and three of these have reference to the empirical work reported here. Proposition 1 (Balanced status); 'If no balanced state exists then forces towards this state will arise. Either the dynamic characters will change, or the unit relations will be changed through action or through cognitive reorganization.' Proposition 2 (Tension states); 'If a change is not possible, the state of imbalance will produce tension.' Proposition 4 (Segregation); 'A balanced state exists if all parts of a unit have the same dynamic character (i.e. if all are positive or all are negative) and if entities with different dynamic character are segregated from each other.' Proposition 4 envisages the possibility of strong boundaries and proposition 1 the possibility of cognitive changes e.g. in the judged distances likely to occur to achieve a state of balance.

The present methods represent some of the types of structures that can be built up within systems that lack balance in this sense. The generalization relating negative feelings to perceptual distortions which preclude the possibility of certain points of view illustrates a type of perceptual blindness. Consider a person (p_1) who likes scholars, heroes and artists in that order but dislikes saints. According to Heider p_1 can conceive of, but will tend to dislike, a person (p_2) who likes saints, artists and heroes in that order but dislikes scholars. How is p_1 to react to a person (p_3) who likes saints, scholars, heroes and artists in that order? To achieve balance p_1 must like p_3 because they both like scholars, but dislike p_3 because he (p_3) likes saints who are disliked by p_1 . People of the type p_1 must find it very difficult to conceive of people of the type p_3 and it is precisely the orders given by these people which are excluded from the configuration of the stimuli as perceived by people of the type p_1 . In fact this configuration excludes all four orders which place saints and scholars together at the beginning of a preference order (see Keats, 1962, p. 358). Similar arguments can be developed to explain examples of perceptual blindness arising in the study of political parties, but here the situation is complicated by more stimuli and a greater incidence of negative feelings. In the case of students' preferences for types of accommodation, the Heider theory of balance accounts for the fact that the subject's own place of residence and the comparison of this with home, give rise to separate dimensional spaces, one for each type of residence.

V. CONCLUSION

The studies reported illustrate the sort of result obtained by the multidimensional analysis of rank orders in which it is assumed that this order can be interpreted as the order of proximity of the subjects to the stimuli. When Torgerson's method of multidimensional analysis was used in conjunction with this method it was found that where there were few strong boundaries, Torgerson's method produced scale values which accounted for the rank orders quite well, but where there were many such boundaries, the Torgerson scale values were affected by the location of the subjects. An analysis of the distance ratings for the whole group might indicate that two dimensions are by no means sufficient to account for the distance values, but when individual differences are reduced by considering only subjects with very similar orders of preference two dimensions might be sufficient, as was the case with the political parties. Examples of perceptual distortion or blindness were discovered in both this case and the study of idealized persons. The data obtained by these procedures were shown to illustrate Heider's theory of balance and lend it some empirical support. The multidimensional analysis of rank orders with the slight extensions used here seems to have particular application to field theory, although it would be a mistake to link a particular method of scaling too closely with a particular psychological theory or vice versa.

REFERENCES

- BERGER, J., COHEN, B. P., SNELL, J. L. and ZELDITCH, M. (1962). *Types of formalization in small-group research*. Boston: Houghton Mifflin Co.
- BENNETT, J. P. and HAYS, W. R. (1960). Multidimensional unfolding: determining the dimensionality of ranked preference data. *Psychometrika*, 25, 27-43.
- COOMBS, C. H. (1952). A theory of psychological scaling. *Engineering Res. Inst. Bull.* No. 34. Ann Arbor: Univer. Michigan Press.
- KEATS, J. A. (1962). Attitudes towards idealized types in Australia and Malaya. *J. soc. Psychol.*, 57, 353-362.
- MCELWAIN, D. W. and KEATS, J. A. (1961). Multidimensional unfolding: some geometrical solutions. *Psychometrika*, 26, 325-332.
- SHEPPARD, R. N. (1962). The Analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27, 125-140.
- SHEPPARD, R. N. (1962). The Analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 27, 219-246.
- THURSTONE, L. L. (1927). The Method of paired comparisons for social values. *J. abnorm. Psychol.*, 21, 384-400.
- TORGERSON, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 19, 401-419.
- TORGERSON, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- TUCKER, L. R. and MESSICK, S. (1963). An individual differences model for multidimensional scaling. *ETS Research Bulletin*, 63-13 Princeton, N.J.: Educational Testing Service.

ITEM ANALYSIS BY PROBIT AND FRACTILE GRAPHICAL METHODS

By RHEA S. DAS

Indian Statistical Institute, Calcutta

This paper considers five basic questions of item analysis and describes an integrated approach for obtaining the answers which utilizes the methods of probit and fractile graphical analysis. Item difficulty is interpreted as the limen at which 50 per cent of the subjects pass; discriminating power is given by the probit regression coefficient; the item-ability relation is specified by the probit regression equation; chi square provides an objective test of whether or not the item meets the normal ogive assumption; and plotting items in terms of the limen and regression coefficient permits the selection of items comparable in terms of ability, difficulty, and discrimination. Techniques for the practical application of this approach are also outlined and illustrated.

I. INTRODUCTION

Classical mental test theory assumes that the abilities underlying performance are continuous and normally distributed variables [19]. It further assumes that the probability that a subject will answer an item correctly is a normal ogive function of his true ability, i.e., the degree to which he possesses the ability underlying test performance [2, 6, 10, 12, 14, 15, 17, 18, 21]. These assumptions provide a theoretical basis for an integrated approach to item analysis, in which refined estimates of parameters to describe the items can be obtained, and from which extensive information can be obtained by the test constructor and analyst. Despite these advantages, currently accepted procedures are of an *ad hoc* nature, and when used in combination, lack logical integration. An alternative approach to item analysis, although based on the same assumptions, uses the probit and fractile graphical methods to obtain more informative, more accurate, and more integrated estimates of item parameters. Two decades ago, Ferguson [6], Lawley [12, 13], and Finney [7] expressed logically the assumptions and rationale underlying this approach. Here, their rationale is extended to cover more problems of item analysis, and emphasis is placed upon practical estimation procedures.

The fundamental principle of probit analysis is to transform an ogive or sigmoidal response curve into a straight line using a transformation of the response probabilities based on the normal integral [8]. As pointed out by Finney, Fechner originally thought of the idea in 1860 while considering weight-discrimination results in psychophysics. The basic terms for the probit analysis of items are as follows. Let P refer to the proportion of testees passing an item and X to a measure of ability, usually the total test score or criterion score. Then the probit Y is the unit normal deviate corresponding to P , plus 5, and is

therefore linearly related to ability, X , by the equation,

$$Y = 5 + \frac{1}{\sigma}(X - \mu), \quad (1)$$

where σ and μ are the parameter values of the standard deviation and mean respectively. This equation will be recognized as the familiar equation for the linear transformation of test scores,

$$Z = K + \frac{S}{\sigma}(X - \mu), \quad (2)$$

where $K = 5$ and $S = 1$. Equation (1) is the probit regression equation, and $b = 1/\sigma$ is the probit regression coefficient.

Building upon these concepts, an integrated approach to item analysis is possible which provides answers to the following five questions:

- (i) How difficult are the items?
- (ii) Do the items distinguish between successful and unsuccessful subjects?
- (iii) In what way are the items related to overall performance?
- (iv) Which items in the test are best?
- (v) How can two tests be set up which are comparable in all ways but have different items?

These questions can be recognized in the context of 'item analysis' as referring respectively to (i) item difficulty, (ii) item discrimination, (iii) item ability relation (iv) item selection, (v) item composition for parallel tests. Each of these questions will now be considered in turn. First difficulty: if the proportion of subjects passing the item is a normal ogive function of ability, the limen or threshold of the item may be given by the mean of the unit normal curve, and hence the value of X which gives the probit $Y = 5$ [7]. The sample mean for an item will be referred to as m . The theoretical and practical advantages of interpreting the limen as an index of difficulty, measured in terms of ability values, were pointed out originally by Ferguson [6] and Lawley [12, 13]. Considering discrimination next, attention may be drawn to the fact that as discriminating power increases, the slope of the regression of item success on ability becomes steeper. Thus when performance on an item is described by probits so that its regression on ability is linear, the probit regression coefficient, b , may be employed as an index of discrimination. For the sample, s is the standard deviation, and $b = 1/s$. Ferguson [6] and Lawley [12] also proposed the use of the standard deviation or its reciprocal as measures of discrimination, but used either the constant process or maximum likelihood estimation procedures. The relation between item performance and ability is described by the probit regression equation (see Figure 1, p. 56).

It is desirable to have some method for estimating the error of measurement of the values obtained by the above procedures. In the present context, the error arises in measuring the trend of item performance in relation to ability. A graphical technique developed by Mahalanobis [16] is appropriate for estimating

error in the measurement of trends. This technique, known as 'fractile graphical analysis', involves a visual comparison of trends observed in two independent subsamples of an original sample. This technique provides a visual estimate of the error of measurement arising in determining the relation between item performance and ability and hence also for the associated difficulty and discrimination measures [3].

It may now be asked whether this approach to item analysis can provide an objective criterion for item selection. According to the initially stated assumptions, the probability that a subject will answer an item correctly is a normal ogive function of his true ability. As ability is inferred from test score, and as test score is assumed to regress linearly upon ability [19], then this probability should be a normal ogive function of test score. To test whether items satisfy this assumption, a test of the "goodness of fit" of the probit line can be employed. Finney [8] has outlined the procedure for testing the significance of the discrepancy between the frequencies expected in terms of theory (the probit line) and the observed frequencies using the chi square test. The hypothesis of "good fit" is accepted if chi square is not significant, and is rejected if chi square is significant. Only those items for which the chi square test is not significant would be acceptable in terms of the original assumption. Since inferences about the test and individual performance are usually made on the basis of classical mental test theory, the items comprising the test must satisfy that theory's assumptions. The chi square test provides the appropriate objective criterion.

A device for choosing items for parallel tests may be suggested. For each item, difficulty, measured by m , and discrimination, measured by b , will have been obtained. If the items are then plotted with difficulty along the abscissa and discriminating power as the ordinate, pairs or triplets of items can be picked out which are matched in terms not only of difficulty and discrimination, but also of ability, because difficulty is given in ability values along the abscissa. This device integrates the estimates of difficulty and discriminating power of an item with the ability required to deal with it, utilizing the assumptions of classical mental test theory.

Currently accepted procedures provide answers to some of the above questions by using indices of difficulty and discrimination. Probably the most widely adopted index of difficulty is the proportion of subjects passing the item; popular indices of discrimination include the correlations of item success with criterion performance, and the comparison of the proportion of successful testees in high and low criterion groups [1, 5, 10, 11]. If proportion of success is taken as the index of item difficulty, it is seen that it provides only a partial answer to question (i) as it does not indicate the relation between difficulty and overall test performance or ability, and also as it is specific to the subject population, rather than to the test itself. Similarly, correlational and comparative indices of discrimination do not permit estimation of performance at various ability levels, and therefore do not completely answer question (ii). It follows

that question (iii) concerning the relation between item success and ability is not satisfactorily answered by these particular indices. Question (iv) also lacks suitable answers, as no objective criteria for item selection based on the item-ability relationship have been developed for customarily employed methods. A method of answering question (v) has been proposed by Gulliksen, in which items are plotted in terms of the proportion of people passing and item-test correlation [11, p. 208]. The relation between item performance and level of ability, although pertinent, is not revealed by this method. These inadequacies do not occur when the probit method is used for item analysis.

II. ESTIMATION PROCEDURES

In this section, practical procedures for the estimation of difficulty, discrimination and the item-ability relation will be outlined and illustrated. Particular attention will be paid to the application of punched card equipment of the Hollerith or IBM type. Appropriate hand tabulation techniques will also be described. Since these procedures can also be carried out on electronic digital computers, additional comment will also be made on the appropriate procedure for programming and processing of the data. The topics dealt with in this section are: (i) obtaining the item frequency tables; (ii) converting frequencies to probits; (iii) drawing the probit line; and (iv) estimating the probit regression equation.

(i) *Obtaining the item frequency tables.*

The range of test scores is divided into seven to ten class intervals and for each interval the total number of subjects is obtained. The number of subjects in each class interval correctly answering each item is then tabulated or counted. To do this on punched card equipment, the item responses and total score for each subject are punched on one card or one set of cards. Item responses may be punched as given on the answer sheet, or as 'right' or 'wrong'. Where mark-sensing cards are used as answer sheets, item responses will be automatically punched by a reproducing punch. The cards for all the subjects comprise the deck of cards. Divide the deck of cards into sub-decks according to class interval and gang-punch an identifying number into the cards of each sub-deck. An electronic statistical machine can then be used to tabulate the item frequencies for each class interval, and also to punch the item frequencies into summary cards (by use of an associated reproducing punch). If this type of machine is not available, a counting sorter or ordinary sorter may be used to obtain the frequencies.

If hand tabulation is necessary, place the answer sheets into groups according to class interval. For each class interval, tabulate the number of right responses for each item, also noting the total number of subjects.

For analysis on an electronic digital computer, item responses must be entered into the computer using the appropriate input medium (cards, tape, etc.) and format. Scoring the responses may be carried out by the computer,

or the total scores may also be entered along with the item responses. By programming, the computer can be instructed to obtain and store the item frequencies in the memory.

(ii) *Converting frequencies to probits.*

Each frequency right is to be converted to a proportion by dividing it by the total number of subjects in its class interval. Transform the proportions to probits using probit tables [7, Table I; 9, Table IX]. These computations are illustrated in Table I.

TABLE I. COMPUTATION OF THE PROBIT VALUES FOR ONE ITEM

Test Score Interval	Score Midpoint	Number of Testees		Proportion Right	Probit Right
(1)	(2)	Total (3)	Right (4)	(5)	(6)
0-4	2	19	3	0.1579	4.01
5-9	7	27	12	0.4444	4.85
10-14	12	31	25	0.8065	5.88
15-19	17	19	16	0.8421	5.99
20-24	22	33	30	0.9091	6.34
25-29	27	81	79	0.9753	7.05
30-34	32	27	27	1.0000	8.09†

† 8.09 taken as probit representing unity

If summary cards giving item frequencies have been prepared in the preceding step, computation of proportions can be done using a calculating punch. Proportions may be transformed to probits by reference to tables after tabulation, or by using a reproducing punch with a master deck of proportion-probit cards. Item success, in probits, should be tabulated for each class interval. Conversion of item frequencies to probits can also be done manually. If an electronic computer is being used, programming would call for computing proportions and reference to a probit table stored in the memory.

(iii) *Drawing the probit line.*

The probit line can be drawn from the data in step ii. Test score is given on the abscissa, and item success in probits along the ordinate as illustrated in Figure 1. One graph should be drawn for each item. Plot each probit against the mid-point of its test score class interval. Following the principle given by Finney [8], draw a straight line which minimizes vertical distances from the plotted points, giving less weight to points above 7.5 and below 2.5 on the ordinate. Figure 1 presents the probit line for the data in Table I. As analytical techniques are more appropriate for electronic computers, including the least squares or maximum likelihood method in the programme would omit this step and directly give the probit regression equation discussed in the next step.

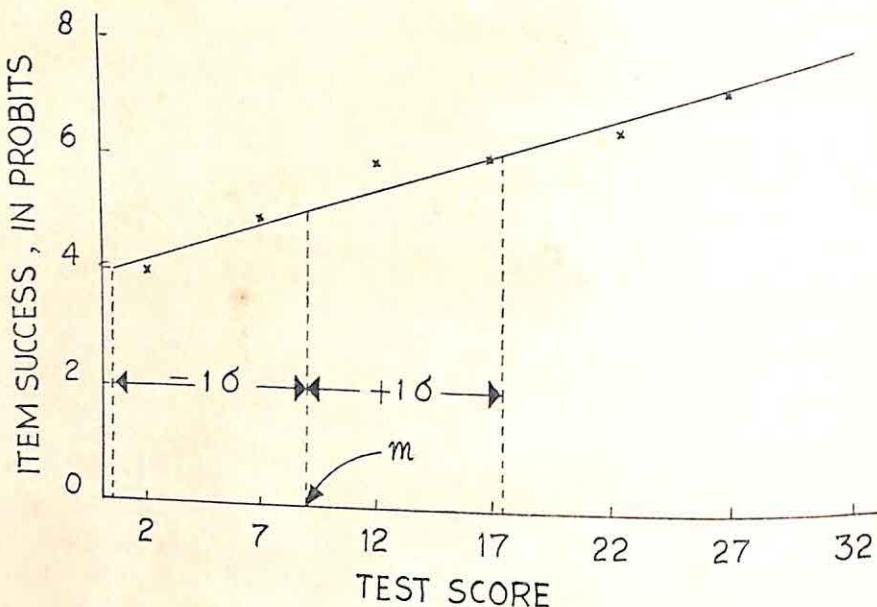


FIGURE 1. Item Success in Probits plotted Against Test Score, giving Probit Line, Mean (m) and Standard Deviation ($\pm 1\sigma$).

(iv) *Estimating the probit regression equation.*

From the probit line, the sample mean, m , and sample standard deviation, s , can be obtained. The value of X corresponding to probit 5 gives m , the difficulty or limen of the item. Subtracting the value of X corresponding to probit 4 from m estimates s , the standard deviation. The reciprocal of s may then be used to give the regression coefficient b , as the index of discrimination. These values are then substituted in equation (1) to give the probit regression equation characterizing the item ability relationship. In Figure 1, m is 9.0, s is 8.5, and hence b is 0.118. Then,

$$Y = 5 + \frac{1}{8.5} X - \frac{1}{8.5} (9.0)$$

$$= 0.118 X + 3.941$$

is the probit regression equation for the item plotted in Figure 1.

III. FRACTILE GRAPHICAL ANALYSIS

Fractile graphical analysis has been suggested in the introduction as a method for visually estimating error of measurement. As this method may not be familiar to psychometric workers, it will be discussed briefly in terms of its original purpose.

Mahalanobis developed fractile graphical analysis as a method for examining the degree of error in estimating trends found in sample surveys [16]. In its original application, two interpenetrating subsamples are drawn randomly with replacement from the original sample. Members of each subsample are placed in decile (or some other fractile) groups on the abscissa. For each group the mean or median ordinate value is obtained, and these ordinate values are plotted and connected for each of the subsamples separately. This procedure is also followed for the combined sample. The latter line gives the sample estimate of the trend, while the area between the lines of the two subsamples gives a graphical estimate of the error.

For item analysis, the procedure may be somewhat modified. Two subsamples are drawn from the set of answer sheets (or cards) and the operations in steps ii and iii above are carried out (see Table II). The probit values for the two subsamples are connected and the area between the resulting lines shaded. The probit values for the combined sample are also connected (see Figure 2). The second line gives the observed values used for the probit analysis, while the shaded area gives a visual estimate of the error associated with the observed

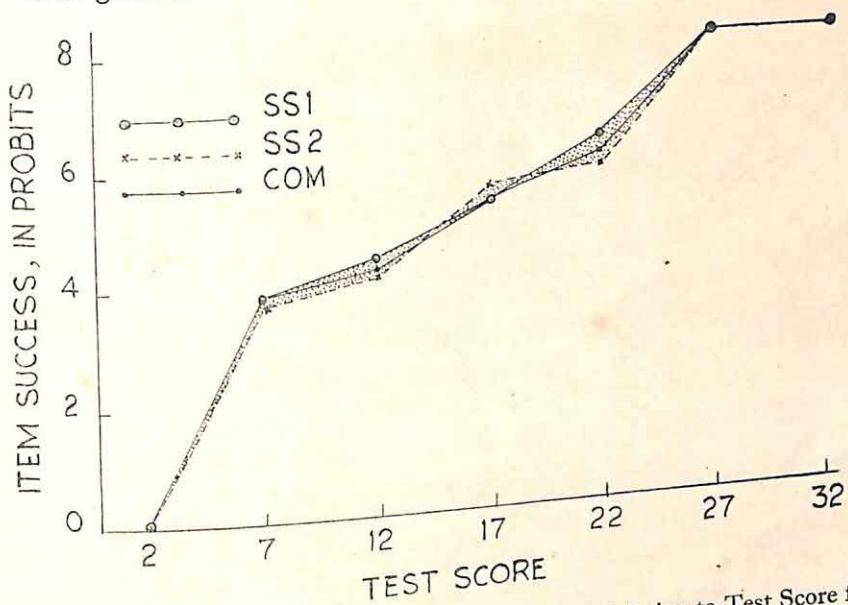


FIGURE 2. Fractile Graphical Analysis of Item Success in Relation to Test Score for Two Subsamples (SS1 & SS2) and Combined Sample (Com).

values themselves. Comparison of the graphs for different items will, in practice, reveal considerable variation in the size of the error area. Greater confidence can be placed in the observed relationship between item success and ability when the error area is relatively small. This method might be employed in the examination and presentation of reliability and validity data, as well as for item analysis data as outlined here [3, 4].

TABLE II. TABLE FOR FRACITILE GRAPHICAL ANALYSIS FOR ONE ITEM

Test Interval	Score Midpoint	Total frequency†			Number Right			Proportion Right			Probit Right		
		SS 1	SS 2	Com	SS 1	SS 2	Com	SS 1	SS 2	Com	SS 1	SS 2	Com
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
0-4	2	8	6	14	0	0	0	0.1429	0.1304	0.1364	3.92	3.87	3.92
5-9	7	21	23	44	3	3	6	0.3077	0.2443	0.2593	4.50	4.19	4.36
10-14	12	13	14	27	4	3	7	0.6000	0.6667	0.6364	5.25	5.44	5.36
15-19	17	5	6	11	3	4	7	0.9167	0.7500	0.8500	6.41	5.67	6.04
20-24	22	12	8	20	11	6	17	1.0000	1.0000	1.0000	8.09‡	8.09‡	8.09‡
25-29	27	16	14	30	16	14	30	1.0000	1.0000	1.0000	8.09‡	8.09‡	8.09‡
30-34	32	7	9	16	7	9	16	1.0000	1.0000	1.0000			

† SS 1 = sub-sample 1
 SS 2 = sub-sample 2
 Com = combined sample.

‡ probit value of 8.09 taken
 to represent unity.

IV. A CRITERION FOR ITEM SELECTION

An objective criterion for item selection has been proposed which can be employed where items have been analyzed according to the procedure outlined above. This criterion tests whether the item satisfies the assumption of a normal ogive relation between probability of item success and test score. Chi square serves as the objective statistical criterion. Those items for which chi square is significant would be rejected and those for which it is not significant would be selected.

The appropriate computational steps are presented in Table III and are

TABLE III. COMPUTATION OF CHI SQUARE FOR ONE ITEM

Test Score Interval	Midpoint	Expected		Total n	Number Right			$\frac{(r-nP)^2}{nP(1-P)}$
		Y	P		Observed r	Expected nP	Discrepancy $r-nP$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0-4	2	4.177	0.2061	19	3	3.92	-0.92	0.272
5-9	7	4.767	0.4090	27	12	11.04	0.96	0.141
10-14	12	5.357	0.6406	31	25	19.86	5.14	3.701
15-19	17	5.947	0.8289	19	16	15.75	0.25	0.023
20-24	22	6.537	0.9382	33	30	30.96	-0.96	0.482
25-29	27	7.127	0.9834	81	79	79.66	-0.66	0.329
30-34	32	7.717	0.9967	27	27	26.91	0.09	0.091
								chi square = 5.039

briefly explained below [8]. They may be followed manually, on punched card equipment, or on an electronic computer.

1. Calculate the expected probit value for each class interval using the probit regression equation (1). The value of X to be substituted in the equation is the mid-point of the respective class interval.

2. Convert the expected probit values to proportions reading from the probit table [8, 9].

3. Determine the expected frequencies, nP , for each class interval by multiplying the number of subjects in it, n , by the expected proportion, P . In Table III, P is given in col. 4, n in col. 5, and nP in col. 7.

4. Subtract expected frequency nP from the observed frequency right,

r. In Table III, col. 6 gives r , col. 7 gives nP , and col. 8 gives $r-nP$.

5. Compute $(r-nP)^2/nP(1-P)$ for each class interval. For Table III this is $(\text{col. 8})^2/(\text{col. 7})(1-\text{col. 4})$.

The sum of these values for all of the class intervals gives chi square with $k-2$ degrees of freedom, where k is the number of class intervals or ability groups. For the illustrative example in Table III, chi square = 5.039 with 5 degrees of freedom, which is not statistically significant thus permitting acceptance of the normal ogive assumption.

Table IV gives chi square values for 33 items of a 40 item test. Of the 33 items, chi square is significant for 14, and not significant for the remaining 19. Those items remaining would thus be selected. It will be noted that seven items in Table IV remain unaccounted for. No probit line could be fitted visually to those items which would give estimates of difficulty and discrimination, because the proportion of success was equal at all levels of ability, resulting in a horizontal probit line. These items were automatically rejected.

TABLE IV. STATISTICS FOR SELECTING AND ORDERING ITEMS OF A 40 ITEM TEST
(based on 237 subjects)

Item Number	Mean	Standard Deviation	Probit Regression Equation	Chi square†
(1)	(2)	(3)	(4)	(5)
1	8.00	16.00	$Y=0.062X+4.500$	13.366‡
2	16.00	7.75	$Y=0.129X+2.935$	8.726
3	13.00	15.00	$Y=0.067X+4.133$	11.296‡
4	16.00	8.00	$Y=0.125X+3.000$	15.093‡
5	16.00	12.00	$Y=0.080X+3.720$	13.675‡
6	27.00	7.50	$Y=0.133X+1.400$	2.566
7	23.00	17.00	$Y=0.059X+3.647$	11.979‡
9	14.00	9.00	$Y=0.111X+3.444$	18.099§
10	9.00	9.00	$Y=0.111X+4.000$	5.039
12	8.00	10.00	$Y=0.100X+4.200$	5.479
13	11.00	10.00	$Y=0.100X+3.900$	4.274
14	14.00	7.00	$Y=0.143X+3.000$	21.159§
15	24.00	12.00	$Y=0.083X+3.000$	9.847
16	20.00	9.50	$Y=0.105X+2.895$	10.802
17	14.00	7.75	$Y=0.129X+3.194$	11.515‡
18	10.00	6.50	$Y=0.154X+3.462$	5.227
20	11.00	10.00	$Y=0.100X+3.900$	16.435§
21	16.00	5.50	$Y=0.182X+2.091$	4.391
22	16.00	14.00	$Y=0.071X+3.857$	12.161‡
23	24.00	9.75	$Y=0.103X+2.538$	23.383§
24	13.00	10.00	$Y=0.100X+3.700$	3.676
25	12.00	10.25	$Y=0.098X+3.829$	9.426
26	21.00	7.00	$Y=0.143X+2.000$	5.707
27	14.00	6.50	$Y=0.154X+2.846$	5.763
29	18.00	8.00	$Y=0.125X+2.750$	4.237
30	13.00	9.00	$Y=0.111X+3.556$	4.339
31	17.00	8.75	$Y=0.114X+3.057$	14.658‡
32	15.00	15.00	$Y=0.067X+4.000$	87.046§
33	17.00	12.00	$Y=0.083X+3.583$	5.445
35	21.00	9.00	$Y=0.111X+2.667$	1.552
36	20.00	7.25	$Y=0.138X+2.241$	3.816
37	20.00	7.50	$Y=0.133X+2.333$	31.457§
40	22.00	11.50	$Y=0.087X+3.087$	3.419

† 5 degrees of freedom. ‡ $P < 0.05$. § $P < 0.01$.

V. A CHART FOR PARALLEL TESTS

In order to pick out items for parallel tests, it would be desirable to select them in terms not only of difficulty and discrimination but also of level of ability. This may be accomplished by plotting items on a chart with difficulty as the limen, m , in ability values, and discrimination as the regression coefficient, b . As an illustration, the 19 items selected from Table IV have been plotted in Figure 3. Difficulty values are taken from col. (2) of Table IV, and discrimination from the first element of the regression equation in col. (4). From Figure 3,

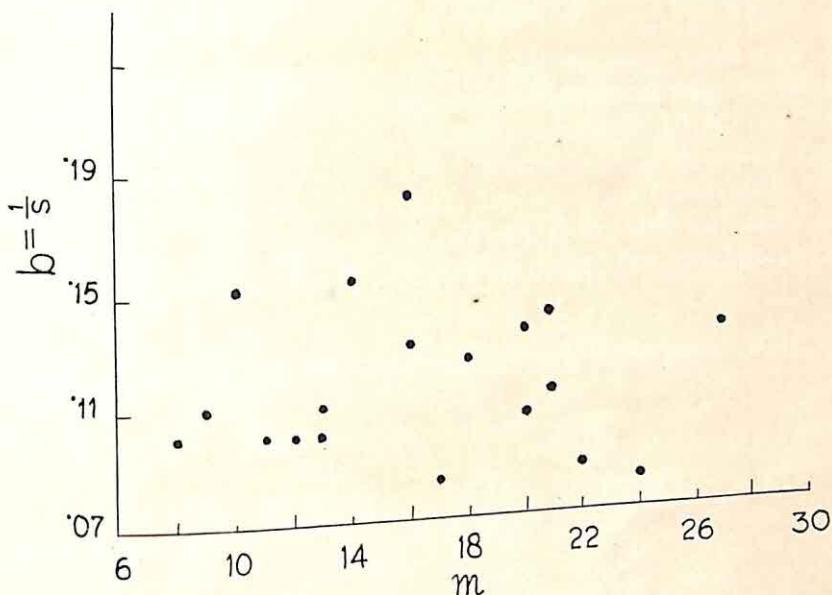


FIGURE 3. Item Selecting Chart for Matching Items in Terms of Mean (m) and Regression (b).

it should be possible to pick out pairs or triplets of items, depending on whether two or three parallel tests are desired. This procedure would assure the test constructor that sets of items picked out in terms of the limen and regression coefficient would be comparable and also that for all items, item success in probits would be linearly related to ability.

This chart could also be used to order items within a test in terms of ability level, difficulty and discrimination. By picking out items from left to right in Figure 3, the 19 items accepted from Table IV can be rearranged in the following order, from least difficult and discriminating to most difficult and discriminating: 12, 10, 18, 13, 25, 24, 30, 27, 2, 21, 33, 29, 16, 36, 35, 26, 40, 15, and 6.

VI. SOME EMPIRICAL OBSERVATIONS

Two questions may be asked about the actual computations of a probit analysis: (i) how reliable is the visually fitted probit line? and (ii), is the probit line affected by the choice of ability groups? These two questions were investigated empirically. The data obtained are summarized below.

Reliability was examined by having the entire set of computations for the test analysis reported in Table IV carried out independently by two different persons. To test whether the two sets of probit lines were giving similar estimates, the limen values were correlated. For the 33 items of Table IV, a product moment correlation of 0.96 was obtained between the two sets of limen values. This result lends confidence to the use of estimates obtained from the visually fitted probit line.

To determine whether the manner of classifying subjects into ability groups would affect the probit line, ability groups were set up using two types of class intervals: (i) equal score intervals, e.g., 0-4, 5-9, 10-14; and (ii) equal frequency, i.e., dividing the score range into such intervals that an equal number of subjects would be allocated to each class interval. The data analysed were those given in their final form in Table IV; the two sets of class intervals and their respective frequencies are given in Table V. For the second set, intervals were

TABLE V. FREQUENCY DISTRIBUTIONS FOR TWO SETS OF CLASS INTERVALS

Set 1		Set 2	
Class Intervals	n	Class Intervals	n
(1)	(2)	(3)	(4)
0-4	19	0-6	35
5-9	27	7-11	29
10-14	31	12-17	27
15-19	19	18-23	27
20-24	33	24-25	27
25-29	81	26-27	29
30-34	27	28-29	36
		30-33	27

chosen so that the number of subjects would be as close to 29 as possible, at the same time allotting all subjects with the same score to the same class interval. Probit lines were drawn for both sets and a *t*-test of the difference between the two sets of limen values was computed. For 33 items, the average difference between the limens was 0.692, and the value of *t* obtained was found not to be significant (*t* = 1.323). The two probit lines for a typical item with a difference of 1 between the two limen values are given in Figure 4. It may be noted that even though the frequency distribution for the first set of class intervals is highly skewed (see col. 2 of Table V), the estimates obtained by the two methods do

not differ significantly. This empirical comparison suggests that the probit line is, to a considerable extent, independent of the basis for selecting class intervals, and hence, frequency and range of ability within groups.

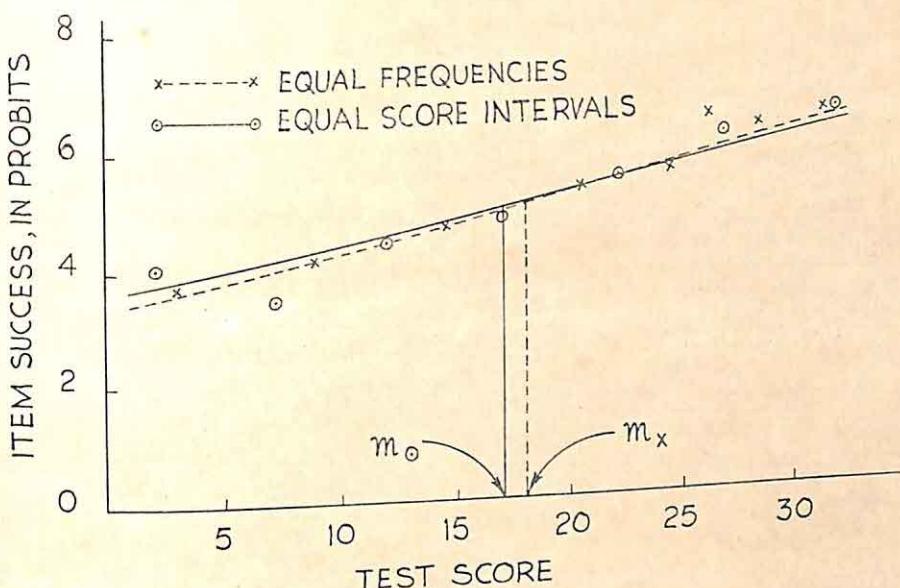


FIGURE 4. Probit Lines and Means Comparing Ability Groups Set Up With Equal Frequencies (m_x) and Equal Score Intervals ($m\ominus$).

REFERENCES

- [1] ADKINS, D. C. (1947). *Construction and Analysis of Achievement Tests*. Washington: U.S. Govt. Printing Office.
- [2] CRONBACH, L. J. and WARRINGTON, W. G. (1952). Efficiency of multiple-choice tests as a function of spread of item difficulties. *Psychometrika*, XVII, 127-147.
- [3] DAS, R. S. (1960). Applications of fractile graphical analysis to psychometry: I. Item analysis. *Psychol. Stud. (India)*, V, 11-18.
- [4] DAS, R. S. and SHARMA, K. N. (1960). Applications of fractile graphical analysis to psychometry: II. Reliability. *Psychol. Stud. (India)*, V, 71-77.
- [5] DAVIS, F. B. (1955). Item Selection Techniques. Chap. 9 in Lindquist, E.F. (ed.) *Educational Measurement*. Washington: American Council on Education.
- [6] FERGUSON, G. A. (1942). Item selection by the constant process. *Psychometrika*, VII, 19-29.
- [7] FINNEY, D. J. (1944). The application of probit analysis to the results of mental tests. *Psychometrika*, IX, 31-39.
- [8] FINNEY, D. J. (1947). *Probit Analysis*. Cambridge: University Press.
- [9] FISHER, R. A. and YATES, F. (1957). *Statistical Tables for Biological, Agricultural and Medical Research*. Fifth edition. London: Oliver and Boyd.
- [10] GUILFORD, J. P. (1936). *Psychometric Methods*. First edition. New York: McGraw-Hill.
- [11] GULLIKSEN, H. (1950). *Theory of Mental Tests*. New York: Wiley.

- [12] LAWLEY, D. N. (1943). On problems connected with item selection and test construction. *Proc. Roy. Soc. Edin.* LXI, 273-287.
- [13] LAWLEY, D. N. (1943). The factorial analysis of multiple item tests. *Proc. Roy. Soc. Edin.* LXII, 74-82.
- [14] LORD, F. M. (1952). A theory of test scores. Psychometric Monograph No. 7.
- [15] LORD, F. M. (1953). An application of confidence intervals and maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, XVIII, 57-76.
- [16] MAHALANOBIS, P. C. (1960). A method of fractile graphical analysis. *Econometrica*, XXVIII, 325-351.
- [17] MOSIER, C. I. (1940). Psychophysics and mental test theory: fundamental postulates and elementary theorems. *Psychol. Rev.*, XLVII, 355-366.
- [18] RICHARDSON, M. W. (1936). The relation between the difficulty and the differential validity of a test. *Psychometrika*, I, 33-49.
- [19] SOLOMON, H. (ed.) (1961). *Studies in Item Analysis and Prediction*. Stanford: Stanford University Press.
- [20] TORGERSON, W. S. (1958). *Theory and Methods of Scaling*. New York: Wiley.
- [21] TUCKER, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika*, XI, 1-13.

PROMAX: A QUICK METHOD FOR ROTATION TO OBLIQUE SIMPLE STRUCTURE

By ALAN E. HENDRICKSON and PAUL OWEN WHITE

Institute of Psychiatry, University of London

A new method for analytical rotation to oblique simple structure is described. Orthogonal simple structure is achieved by means of any of several existing rotation methods and this is then transformed into an oblique solution.

I. INTRODUCTION

During recent years, numerous attempts have been made to develop methods for the analytic determination of oblique simple structure (see Warburton, 1963). From the first attempt by Carroll (1953) to the more recent work of Kaiser and Dickman (1959) with their Binormamin criterion, these methods have become increasingly successful in that they approximate ideal graphical solutions more closely. This success, however, has been won at the price of increasing computational time to the point where the desired oblique solution is often beyond the means of a research worker.

This paper describes a general method for quick rotation to oblique simple structure. Although all the computations have been carried out with the aid of a high speed computer, the solutions could easily have been obtained by hand computation. Furthermore, for the particular data that have been considered, the results appear to be at least as good as those obtained by any existing analytic methods.

II. RATIONALE

Previous oblique analytic methods have been iterative procedures that have relaxed the restriction to orthogonality throughout the computations. Rather than attack the problem in this fashion with a new criterion, it was decided to use an orthogonal method first, construct an "ideal" oblique solution from the results of the orthogonal rotation, and then rotate the orthogonal results to a least squares fit to this ideal solution. The now very well known and readily available Varimax method of Kaiser (1956, 1958) was used as the orthogonal criterion. This choice was dictated by two major considerations; firstly, Varimax is relatively fast as an iterative method, and secondly, the results which Varimax typically gives were well known.

III. THE CONSTRUCTION OF THE IDEAL SOLUTION

Starting with a matrix of factor loadings that has been rotated to orthogonal simple structure, the problem was to construct an ideal, or "pattern" matrix as a function of the orthogonal solution. Perhaps the most simple pattern matrix

is generated by inserting ± 1.0 for salient factor loadings and putting 0's elsewhere.

However, whenever the least squares fit of the orthogonal solution to this simple pattern matrix was computed the result seldom seemed satisfactory. The solution obtained from such a procedure, though commonly quite oblique, was very far from what most people would call simple structure. The trouble seemed to lie in the definition of 'salient'. For example, a cut-off point of 0.20 or above will usually put far too many 1's in the matrix and a cut-off point as high as 0.50 resulted in variables with loadings of 0.49 not being given sufficient weight.

There was a need for differential weighting of the variables, and the logical conclusion of this line of argument is to have an infinite set of weights corresponding to an infinite number of possible loadings. Now the problem was to find a set of weights which would correspond to our intuitive idea of how much weight should be given to a given factor loading. The accepted ideal solution is that high loadings should be increased while low loadings should be made smaller; weights for high loadings should therefore be proportionately larger than those given low loadings. A family of sets that meet this intuitive notion of greater proportionate weights for high loadings are those sets that consist of powers greater than 1 of the corresponding loadings. It was therefore proposed to generate a matrix wherein each element would be a function of the orthogonal loading, the length of the test vector, and the length of the factor vector.

We define a matrix $P = (p_{ij})$, such that:

$$p_{ij} = |a_{ij}^{k+1}| / a_{ij}, \quad (1)$$

with $k > 1$. Each element of this matrix is, except for sign which remains unchanged, the k th power of the corresponding element in the row-column normalized orthogonal matrix. Then find the least squares fit of the orthogonal matrix of factor loadings to the pattern matrix generated by equation (1):

$$L = (F' F)^{-1} F' P, \quad (2)$$

where L is the unnormalized transformation matrix of the reference vector structure, F is the orthogonal rotated matrix, and P is the matrix derived from the orthogonal matrix defined in eqn. (1). This equation is the "Procrustes" equation recently described by Hurley and Cattell (1962). The columns of L are normalized such that their sums of squares are equal to unity. This provides the transformation matrix from the orthogonal factors to the oblique reference vectors. From this point standard and well-known formulae (Harman, 1960) can be used to compute the intercorrelations among the oblique primary factors and the matrix of primary factor loadings.

The function (1) was selected as a means of increasing the ratio of high to low loadings in the 'pattern' matrix P , thus in effect forcing the smaller loadings into the hyperplane without having to decide arbitrarily by inspection on a cut-off point between the high loadings or salients and the low or 'hyperplane' loadings. It was hoped that the orthogonal solution would be sufficiently

close to the desired oblique solution so that successive powering of the Varimax loadings would force the relatively low loadings into the hyperplane, thus yielding a 'pattern' matrix P towards which the Procrustes procedure indicated in eqn. (2) would rotate the orthogonal solution. For a given matrix of orthogonal factor loadings, normalized by rows and by columns, eqn. (1) yields a whole family of matrices P , one for each value of k , any one of which might be used for the Procrustes rotation in eqn. (2). The name PROMAX is suggested for the technique and the class of solutions suggested in this paper.

IV. CHOICE OF A POWER

The question now arises as to what value should be given to k . This is analogous to the problem of deciding on a value for the weighting parameter in Carroll's Oblimin solution.

TABLE I. COMPARISON OF PROMAX SOLUTION WITH VISUAL SOLUTION OF HOLZINGER AND HARMAN FOR THE TWENTY-FOUR PSYCHOLOGICAL TESTS

Test	Primary Factor Pattern Matrices *				Promax Solution $k=4$			
	Holzinger and Harman				75	-07	05	-01
01	70	-07	09	04	49	-03	-03	00
02	46	-02	-01	02	63	02	-12	-05
03	60	02	-08	-01	61	07	-03	-09
04	58	07	00	-05	-01	81	10	-07
05	-04	83	09	-06	01	80	-05	06
06	-04	83	-08	09	-02	91	05	-13
07	-05	93	04	-12	25	50	16	-09
08	22	52	16	-08	-01	87	-17	13
09	-08	90	-20	16	-32	07	80	11
10	-32	08	78	08	-15	04	62	25
11	-18	07	61	24	13	-15	77	-10
12	14	-15	79	-12	34	02	61	-20
13	35	01	64	-20	-15	14	05	54
14	-24	18	01	58	02	00	-07	56
15	-07	03	-10	61	39	-11	-07	41
16	31	-08	-07	47	-14	01	05	72
17	-24	04	01	76	24	-23	13	54
18	14	-19	13	59	16	00	03	38
19	08	02	03	42	43	24	-05	12
20	37	25	-05	17	37	-03	31	11
21	32	-01	32	13	34	25	-15	29
22	26	28	-15	34	54	20	06	04
23	49	21	08	08	02	24	38	21
24	-01	27	39	18				

Primary Factor Correlations *

100	58	100			100	58	100	
	46	46	100	100		52	44	100
	58	51	60	100		55	50	54
								100

* Decimal Points Omitted

In the absence of any clear rationale for selecting a particular value of k , it was decided to select data with both a known factor structure and an oblique rotation solution which is generally accepted as adequate and to apply the procedure with several values of k in the hope that an approximation to the adequate oblique simple structure would be found. It was argued that even though it might be necessary to try several values of k before an adequate solution was found, this would be an economically feasible procedure.

TABLE II. COMPARISON OF PROMAX SOLUTION WITH VISUAL SOLUTION OF CATTELL AND DICKMAN FOR THEIR BALL PROBLEM

Primary Factor Pattern Matrices *

Measure	Cattell's Visual Solution					Promax Solution $k=4$			
	01	104	-06	00	02	105	-06	01	08
02	-05	102	-03	01		10	92	-03	-03
03	12	-29	88	-05		07	-25	88	-04
04	05	-05	01	100		05	-05	01	101
05	-84	-10	00	03		-87	-08	-01	-01
06	108	-10	-01	01		107	-09	00	07
07	101	-03	01	02		102	-03	02	08
08	-87	-06	01	03		-89	-05	01	-02
09	92	-03	05	03		93	-03	06	08
10	-05	102	-03	01		10	92	-03	-03
11	-50	122	-02	-04		-33	110	-01	-11
12	12	83	05	-02		24	75	06	-05
13	05	94	-04	02		19	85	-03	-01
14	03	-101	05	-01		-12	-90	04	03
15	02	-15	78	-03		-02	-13	79	-03
16	15	-29	87	-02		10	-25	87	-01
17	-54	-31	64	-05		-60	-27	63	-07
18	-03	00	93	06		-05	01	93	05
19	-24	28	91	02		-20	26	92	-01
20	-05	-02	01	99		-05	-02	01	99
21	-07	00	01	98		-07	-01	02	98
22	-03	-02	02	100		-02	-03	02	99
23	-13	05	01	-101		-13	05	01	-102
24	-10	-08	03	-95		-12	-07	03	-95
25	10	39	56	-09		16	36	57	-10
26	17	05	86	-03		18	05	86	-02
27	59	41	-10	03		66	37	-09	05
28	59	08	57	-02		60	08	57	01
29	-02	13	85	10		00	13	85	09
30	-60	-08	02	59		-62	-07	01	55
31	82	-13	00	-43		81	-13	00	-37
32	08	84	16	31		21	76	17	29

Primary Factor Correlations *

100			100		
76	100		68	100	
21	-05	100	20	-10	100
-11	-06	03	100	-16	00
				02	100

* Decimal points omitted.

Promax was first tried on the now classic data from the Twenty-Four Psychological Tests of Holzinger and Harman (1960). The results were very encouraging. With $k=4$, the results seemed to be better than the already quite adequate Binormamin provided by Kaiser and Dickman (1959) and was indeed somewhat cleaner than the subjective visual solution which we regarded as a criterion. Table I presents a comparison of the Holzinger and Harman visual solution with the Promax solution with k equal to 4.

To test the generality of 4 as an adequate power, we used the same procedure on other data. Table II presents the analysis of Cattell and Dickman's (1962) ball problem. The Promax solution with $k=4$ was compared with Cattell's visual rotation which was based on the Binormamin solution as a starting point. It would be superfluous to present the Binormamin solution here for the simple reason that the maximum difference between any one element of the Binormamin matrix and the corresponding element of the Promax matrix was 0.04, and the large majority of elements differ by only 0.01 or not at all. The hand rotation differs little from ours, and in view of the very close correspondence between Promax and Binormamin, it seems almost certain that Cattell would have

TABLE III. COMPARISON OF THURSTONE'S SOLUTION TO HIS BOX PROBLEM AND PROMAX WITH $k=2$

Measure	Primary Factor Pattern Matrices *			Promax Solution $k=2$	-02	-05
	Thurstone's Solution					
01	100	00	00	101	-02	-05
02	03	99	01	00	100	00
03	01	02	98	-05	-02	101
04	49	76	00	47	76	-03
05	40	01	87	35	-05	87
06	01	41	82	-05	38	83
07	74	54	-03	74	53	-06
08	87	02	40	87	-02	37
09	-01	80	46	-06	79	46
10	63	66	-01	62	65	-04
11	69	01	66	66	-04	64
12	-01	64	65	-06	62	66
13	98	04	-01	100	01	-06
14	-02	97	06	-06	98	06
15	03	-02	96	-03	-07	99
16	29	38	72	24	34	73
17	66	48	28	64	46	26
18	97	-03	01	99	-06	-03
19	07	95	-04	05	96	-05
20	00	03	98	-05	-01	100

Primary Factor Correlations *

100			100		
23	100		29	100	
10	22	100	22	29	100

* Decimal Points Omitted.

achieved the same result had he started from the Promax solution. It is of interest to note that, using ± 0.10 as a criterion, the visual rotation improved the hyperplane count on only one factor.

Table III compares Thurstone's (1947) solution to his Box Problem and the Promax solution with $k=2$. The powers higher than 2 produced solutions similar to other analytic solutions and shared the same defect; i.e. they were too oblique. However, the Promax solution shown seems to be almost as good as the visual solution provided by Thurstone. Although the Promax hyperplane count (± 0.01) is identical to that of Thurstone's visual solution, the latter is obviously just slightly 'cleaner' than the analytic solution. The Promax solution with $k=3$ is virtually identical to the Binormamin solution; the maximum difference between any two corresponding elements of the matrices being less than 0.02, and the majority being identical to two decimal places.

In addition to the three textbook examples presented here, Promax has since been tried on many sets of data from studies currently in progress. The general conclusion to be drawn is that for the majority of cases, the optimal value for k is 4; however, for the occasional factor analysis where the data is particularly 'cleanly' structured, a lower power seems to provide the best solution.

REFERENCES

- CARROLL, J. B. (1953). Analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18, 23-38.
- CARROLL, J. B. (1958). Solution of the oblimin criterion for oblique rotation in factor analysis. Unpublished manuscript.
- CATTELL, R. B. and DICKMAN, K. W. (1962). A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychol. Bull.*, 59, 389-400.
- DICKMAN, K. W. (1960). The factorial validity of a rating instrument. Unpublished doctoral dissertation, University of Illinois.
- HARMAN, H. H. (1960). *Modern Factor Analysis*. Chicago: University of Chicago Press.
- HURLEY, J. R. and CATTELL, R. B. (1962). The Procrustes program: producing direct rotation to test a hypothesized factor structure, *Behav. Sci.*, 7, 258-262.
- KAISER, H. F. (1956). The varimax method of factor analysis. Unpublished doctoral dissertation, University of California.
- KAISER, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- KAISER, H. F. and DICKMAN, K. W. (1959). Analytical determination of common factors. Unpublished manuscript.
- THURSTONE, L. L. (1947). *Multiple Factor Analysis*. Chicago: University of Chicago Press.
- WARBURTON, F. W. (1963). Analytical Methods of factor rotation. *Brit. J. stat. Psychol.*, XVI, 165-174.

NOTES AND CORRESPONDENCE

WHAT IS CONSCIOUSNESS?

By J. G. TAYLOR

Professor Burt's plea (this *Journal*, XIV, 145-170) for the restoration of the concept of mind to psychology calls for a vigorous reply. If he had pleaded for a resumption of the systematic study of conscious phenomena, without prejudice to the scientific search for principles to explain them, he would have met with general approval. But that is not what he asks for. His contention is that to account for the phenomena of consciousness it is necessary to postulate an immaterial 'mind' that is independent of the body and has certain dispositional properties, of which the most distinctive is a "capacity for consciousness in the narrower sense of the term, i.e., for cognitive 'awareness'". It is this capacity that makes the postulated mind distinctively 'psychic', i.e., immaterial. Cognitive awareness is further described as "a simple and unanalysable dyadic 'relation'" between the 'mind' and some object, which need not be material, since there can be awareness of one's own conscious states.

If we accept this notion, it is clear that we are committing ourselves in advance to a severe restraint on the scope of scientific enquiry. If we say that consciousness is the product of something that has a capacity for consciousness, we are saying nothing of any moment; but at least it is implied that there is something to look for, something whose properties may explain the capacity in question. But if we say that this something is psychic or immaterial we are locking the laboratory door and throwing away the key. For who can search for something that has no material existence? In what follows I propose to show that awareness of even such an apparently simple thing as the position of a material object is neither simple nor unanalysable, that the concept of 'mind' is unnecessary for its explanation and is indeed incapable of explaining it, but that it can, on the contrary, be fully accounted for in behavioural terms.

Consider my awareness of the unchanging position of the ink bottle on my writing desk. As Henri Poincaré pointed out 60 years ago, in *La Science et l'hypothèse*, the only real sense in which this position may be said to be unchanging is that the same bodily movement always serves to establish contact with it. That is, as long as I remain seated, with my back resting on the back of the chair, the coordinates of the ink bottle, referred to a frame of reference anchored in my body with the origin in, say, the atlas vertebra, are invariant. But these coordinates are not directly represented in the afferent system, and there is therefore no obviously simple way of accounting for the constancy of the perceived position.

Apparently the only afferents from which the position of the bottle can be derived are those depending on the coordinates of its images on the two retinas. But these coordinates are constantly changing with movements of the eyes and head, so that they alone cannot yield the invariant position. Let us consider this as a problem in coordinate geometry. A geometrician, if given the coordinates of the retinal images at any moment and asked to compute the coordinates of the object on a reference frame anchored in my body, would require to know in addition (a) the coordinates of my eyes on the bodily reference frame when my head is in a standard position, (b) the lateral displacement of each eye relative to the median plane of my head, (c) the vertical displacement of my eyes relative to a horizontal plane through my head, and (d) a transformation specifying the momentary orientation of the head relative to the bodily coordinate system. The information under (a) constitutes the parameters of the system, while (b), (c) and (d) are variables. Armed with this information the geometrician can compute the position of the ink bottle, and he will return the same answer no matter what values are assigned to variables (b), (c) and (d), since these, in conjunction with the position of the object, determine the coordinates of the retinal images.

All this information is required for an exact solution, and if one or other of the variable values is not disclosed, the geometrician can only specify a region within which the object is to be found. Within that region its position is indeterminate. Now if my awareness of the position of the bottle is an act of an immaterial 'mind', it is evident that it must utilise all the information that the geometrician requires, or at least physiological analogues of that information, since the perceived position of the bottle remains extraordinarily stable as long as I am in good health and wide awake. It is obviously required in any case, whether there is a 'mind' or not. I shall later show how it is utilised by an organism that has no 'mind'; but for the moment I am concerned with the implications of this requirement for Professor Burt's theory.

The geometrician gets his information in numerical form and he can feed the data into a digital computer and get the answer in a very short time. But the data on which the 'mind' has to operate are in the form of analogues of the numerical data, and very awkward analogues too, according to neurological findings, since many hundreds or even thousands of neurons are involved in representing any one value of each variable, and two values may excite many neurons in common. The computational task which the 'mind' has to undertake is evidently a very formidable one, as it is required to arrive at the answer in a fraction of a second and often more than once in a second. And because it is, *ex hypothesi*, an immaterial entity, it is obliged to process this complicated mass of data without benefit of the hardware built into a material computer. But that is not all. It is not only the ink bottle that has a constant position in my awareness. There are other objects in the environment, each with its distinctive and fixed position, and I am simultaneously aware of all those positions as being constant. If there is a 'mind' that is responsible for all this, its task is not merely formidable, it is forbidding. The onus is on Professor Burt to show how it is accomplished; and he must note that there is no afferent representation of the parameters. He must show how the 'mind' acquires its knowledge of these parameters. This is a vital question, and science will not be content with a vague answer to it.

Professor Burt can no doubt afford to ignore these difficulties while the only rival theories in the field are so ill conceived that he can make logical sport of them, thereby diverting his readers' attention from the much more serious deficiencies of his own theory. Moreover, he blinds himself to the difficulties by committing himself to the declaration that the relation of awareness is "simple and unanalysable".

Let us for a moment concede this last proposition and investigate the implications of Professor Burt's further statement that 'mind' is to be conceived as being a field, analogous to a gravitational or magnetic field. At any point a field force has both magnitude and direction, and for any one point there is but one magnitude and one direction. Superficially there would seem to be nothing in the conscious experience of normal healthy people that is inconsistent with this notion. But scientific curiosity is never content to accept the surface appearance of things as telling the whole truth about them. It insists on peering under the surface, and the most powerful instrument for this purpose is the experimental manipulation of what the scientist guesses to be the independent variables. Most of the guesses are wrong, but sometimes the truth is revealed as an accidental by-product of experimentation.

An obvious way to experiment on the visual perception of space is to distort the ocular input in some way. Surgical removal of cataracts is one way, but there are many ways of producing what Dr. Ashby (in a private communication) described as a reversible lesion whose effects on the input variables are exactly known. If the experimenter is content to describe the immediate perceptual effects of his distorting device, he is not likely to discover much that is new. It is only when the subject is required not merely to describe what he perceives but to live for days or weeks in the distorted visual world that important discoveries are made. The most striking of these is that the perception mediated by the distorted input changes in a variety of ways but always approximating more and more closely to the perception mediated by the normal input before the beginning of the experiment.

The change is always gradual; but in one type of experiment there is a sudden switching from a false perceptual structure to the true structure, and the gradualness takes the form that periods of veridical perception increase in frequency and duration. This happens when the subject wears spectacles in the form of total reflecting prisms, which reverse the ordering of the retinal image along any axis we may choose. If we choose the horizontal axis right and left are perceived in reverse, but after several days during which the spectacles are worn continuously, or at least for a substantial portion of each day, the subject may report a brief period of veridical perception. If we adopt Professor Burt's theory we might say that the field has been reversed in sign along one of its dimensions, and that this reversal becomes more frequent until finally it maintains the new direction continuously. The 'mind' has, so to speak, restructured itself.

There is, however, a very formidable obstacle to this interpretation. To make this clear it will be useful to adopt an arbitrary definition in terms of Professor Burt's theory. Consider a number of points, A, B, C . . . Y, Z, strung along a horizontal line in both objective and perceived space, such that A is to the right of B, B to the right of C, and so on. We perceive these points as being an ordered series. Now assuming that this perceived ordering is the work of a 'mind' having the properties of a field, we may presumably say that the set of objective points is mapped into the space in which the field operates, and that the field strengths at these points are $\alpha, \beta, \gamma, \dots, \psi, \omega$. Our arbitrary definition states that

$$\alpha < \beta < \gamma < \dots < \psi < \omega;$$

and we postulate that the perceived right-left ordering is simply and unanalysably derived from the ordering of the series α to ω .

If the subject puts on spectacles that reverse right and left the objective set of points is mapped in the reverse order into the space of the mental field, so that point A encounters a field force of ω instead of α , B gets ψ instead of β , and so on. But by definition $\omega > \psi > \chi > \dots > \beta > \alpha$, and therefore by hypothesis the external scene is perceived in reverse. Next, to account for the results of prolonged wearing of the spectacles, we may postulate that the psychic field, by some mystical alchemy whose nature I shall not try to guess, gets its potential gradient reversed along one of its dimensions. At first it cannot hold this reversal for more than a second or two at a time, but gradually it gets established as part of the enduring structure of the 'mind'. The mapping of the objective points into the mental space is reversed as before, but now $\omega < \psi < \chi \dots < \beta < \alpha$; and consequently by hypothesis A is perceived as being to the right of B, and so on. Veridical perception is restored.

We are now ready to introduce the obstacle. In two experiments with right-left reversing spectacles performed in Innsbruck under the direction of Professor Erisman and Professor Ivo Kohler both subjects reported that after veridical perception of the spatial ordering of the environment had been stabilised, printed matter continued to be perceived in reverse, even when the spaces containing the print were seen in their correct places. Thus a building on the right hand side of the street was seen in that position, and its various parts were seen in their correct relations to one another, but an inscription on a wall was seen as mirrored. But this implies that the field has suffered a local and limited change of structure such that in place of a smooth gradient of potential we have a gradient that goes smoothly for a bit, then leaps to a markedly different level, continues with a slope of reversed sign, and after another violent leap resumes the original course of the curve. Instead of $\alpha > \beta > \gamma > \delta > \epsilon > \zeta > \eta$ we have something like $\alpha > \beta > \gamma < \delta < \epsilon > \zeta > \eta$, and the interval between β and ζ is exactly the same in both cases, implying that there is a sharp change in slope from β to γ and again from ϵ to ζ . One portion of the field has reverted to its original structure, while the rest retains the new structure. I must leave it to Professor Burt to tell us what sort of a field is capable of changing its spots in this odd fashion. And here I must remind him of his contention that the structure of the field is entirely independent of the objects or existents that are controlled by it. He is thus precluded by his own dogma from

postulating that printed matter causes the field to alter its structure. Indeed he is precluded from accepting my suggestion that the field can alter its own structure by reversing the sign of the gradient along one of its dimensions. According to his theory the reported righting of the reversed perceptual field is an impossibility.

I now propose to use Occam's razor to lop off the *ens* called 'mind' on the ground that it is not necessary for the explanation of consciousness. This will take the form of a demonstration that the phenomena of consciousness can be exactly accounted for in terms of concepts that form part of the currency of natural scientific discourse. For this purpose let us go back to the ink bottle on my desk.

We have seen that the position of the bottle cannot be perceived as constant unless analogues not only of the positions of the retinal images but also of the directions of the eyes in their sockets and of the orientation of the head on the shoulders are registered in the afferent system. But these data in themselves are not enough. There must also be an operation by which they are made to yield the coordinates of the bottle on my bodily reference frame, or at least a satisfactory analogue of these coordinates. An obvious way in which this can be done is for the complex pattern of afferent events to be conditioned as a whole to the motor response of stretching out my hand to grasp the bottle. But the direction and extent of this movement varies with the initial position of my hand, and it follows that an exact (or sufficiently exact) response cannot be determined by conditioning unless the conditioned stimulus includes, in addition to the components already mentioned, proprioceptive afferents as determined by the initial position of the hand. This means that the conditioned stimulus must be conceived as having no fewer than eleven dimensions.

The number of possible patterns of stimulation involving the ink bottle is colossal; but it should be noted that its position is fixed, so that there are only eight degrees of freedom in the choice of patterns. In other words, the patterns involving the bottle may be said to be confined to an 8-dimensional subspace of the 11-dimensional afferent space. This subspace is further subdivided into a number of 5-dimensional subspaces, each associated with a different initial position of the hand. As soon as a point in one of these 5-spaces is conditioned to a response that establishes contact with the object, stimulus generalization ensures that other points in the same subspace will suffice to evoke the same response.

The details of the process are expounded at some length in my book *The Behavioral Basis of Perception* (Yale University Press, 1962), and I shall not repeat them here. It is enough to say that in the first six months of his life a normal infant acquires conditioned reaching responses to objects occupying all possible positions within reach of his hands. The further developments made necessary by increasing mobility are too complex to be described here.

This conditioning implies that a large number of conditioned reflex arcs, or *engrams* as I prefer to call them, are established in the cortex, and I assume that they are all physically distinct from one another, although they may share many neurons in common. Since the brain has more than 10^{10} neurons it is obvious that the number of combinations of anything from a few hundreds of neurons at a time to some tens of thousands or more is big enough to accommodate all the multitudinous memories and skills of even the most gifted of individuals, and there is no need to assume that the cellular molecules may be the carriers of those skills. It is true that Lashley failed to find any direct evidence for the existence of engrams, but his failure sprang from two errors. Firstly, he did not recognize that there are thousands or even millions of engrams mediating responses that all terminate in the same way, so that the destruction of some of them does not prevent a goal from being reached. And secondly, he did not recognize that any one engram involves neurons in widely separated parts of the cortex. A search in a restricted locality for a structure that is in fact a network spread over a much wider area must necessarily fail.

I now assume that an illuminated environment containing many objects can simultaneously activate a number of engrams mediating conditioned responses that are adapted to the positions of those objects, so that the organism may be described as being in a state

of simultaneous readiness for a large number of learned responses, although only one of them can be executed at any moment of time. To give this state of multiple simultaneous readiness a precise definition would require too much space. The full definition will be found in my book.

I finally postulate that this state of simultaneous readiness is identical with visual perception of the spatial ordering of the environment, that is to say, with a limited aspect of consciousness. There are other sectors of behaviour, directed to other aspects of the environment, and readinesses for those forms of behaviour constitute other dimensions of consciousness. The state of readiness is defined without any reference to consciousness, and all the concepts used are fully acceptable to the other biological sciences. Nevertheless it is possible to deduce from the definition all the properties of the visual perception of space, including the constancies and their limitations, what Professor Gibson calls the unbounded character of the visual world, the geometric illusions, and many more. So far the theory has been worked out in detail only within the realm of visual perception, including the perception of colour; but I am convinced that it is generally applicable to the whole field of consciousness.

I wish to emphasize that my theory is not a variant of the double aspect theory. I do not say that consciousness and multiple simultaneous readiness are different aspects of the same thing: I say that they *are one and the same thing*. Nor has the theory anything in common with epiphenomenalism, according to which consciousness is a mere surface decoration that performs no useful function in the organic economy. If the organism were, like a radar-controlled gun, compelled to pursue one line of behaviour to the total exclusion of everything else that might conceivably call for action, few animals would survive long enough to reproduce their kind. But when an animal has the capacity for multiple readiness, *alias* consciousness, it can at a moment's notice switch its ongoing activity to another goal and thereby avert disaster or secure some other advantage. Consciousness, *alias* multiple readiness, is not a useless decoration but an urgent necessity for any complex organism.

The theory not only accounts for normal consciousness but can also explain fully the abnormalities of consciousness that are induced by experimental manipulation of the independent variables. But before illustrating this I must explain that the independent variables are not the afferent impulses alone, but those impulses in conjunction with the behaviour that has been conditioned to them. In the case of the exteroceptors the afferent impulses are joint functions of certain properties of the external environment and the orientation of the receptor organ relative to the body, with the structure of the receptor and the afferent nerves providing the parameters of the system. Because they are *joint* functions they cannot in themselves yield knowledge of the environmental independent variables, but if, on the other hand, they were not single-valued functions, no knowledge of the environment would be possible at all. The values of the exteroceptive functions are, however, uniquely determined by the environment as soon as specific values are assigned to the orientation variables, and if there are single-valued afferent (proprioceptive) functions of the latter, the total afferent pattern is unique and can therefore be uniquely conditioned to a response directed to the relevant environmental object, which provides the reinforcement. The activity of engrams mediating such responses constitutes knowledge of the environment.

It should now be clear that if a transformation is applied to one of the exteroceptive channels, this will result in erroneous knowledge of the environment and, when action is called for, in misdirected responses. But these responses are immediately subjected to negative reinforcement, and the subject at once begins to acquire responses that are correctly adapted to the relevant features of the environment. We may then predict that to the extent that his behavioural errors are corrected, his erroneous knowledge of the environment will be replaced by true knowledge. Normal perception will be restored. On the other hand, if there are features of the environment to which no action is directed, and there is therefore no opportunity for the correction of behavioural errors, erroneous perception of those features will persist. It is this that accounts for the persistence of reversed

perception when the external stimulus takes the form of printed matter, in the experiments referred to above. The subjects had first to solve the difficult problem of finding their way around easily in a mirrored world before they could devote much energy to reading. Actually there is another important factor here, but I shall not take space to discuss it.

The principles are more clearly illustrated by an experiment in which I acted as subject, with the collaboration of Professor Kohler. I wore spectacles in the form of 20° wedge prisms with the bases to the left. One of the immediate effects of those spectacles is that horizontal plane surfaces below the level of the eyes are perceived as sloping down from right to left with an increasing gradient. Two surfaces of this kind were particularly important to me, viz. the ground on which I walked and the work table at which I sat. Behavioural errors induced by the former included aiming my walking stick, carried in my left hand, at a point some distance below the actual level of the ground, with consequent jolts to my joints, and aiming my (defective) right foot at a point slightly above the ground, with a consequent threat to my equilibrium. The errors in response to the work table, or the breakfast table, involved a different motor system, that of the hands and arms. There was no threat to equilibrium, and no responses relative to the gravitational axis needed correction. I had to learn to move my hands along a plane surface instead of a curved surface.

The spectacles were worn for four and a half hours on each of thirteen consecutive days, with the result that there were none of the perceptual after-effects commonly reported at the end of experiments of this kind when the spectacles are worn continuously. This enabled me to make a direct comparison of my perception at the end of the experiment with what it had been at the beginning. By looking through the spectacles with bases to the right I got exactly the same kind of distortion as with bases to the left, except that the errors were reversed in sign. When I stood or walked, this comparison showed that whereas at the beginning the ground appeared to slope down continuously from left to right with an increasing gradient (a series of cross sections through the ground at different distances from my feet would have shown something like a family of exponential curves with the constant multiplier in the exponent tending to zero as distance increased), at the end of the experiment the ground was perceived as having a level path about a yard wide stretching ahead of me from my feet. To the right of this path there was a gentle upward slope, and to the left a steeper downward slope. When I sat at the table its surface appeared as an oblique plane, the difference between the levels at the two sides being approximately the same at the end of the experiment as at the beginning. There was no trace of curvature on any part of the surface, and there was no level section in the middle. Thus two very similar patterns of retinal stimulation produced two strikingly different perceptual structures according as I stood or sat.

If cognitive awareness is the work of an immaterial 'mind' having an independent structure of its own, these results are incomprehensible. If, on the other hand, consciousness is a state of multiple simultaneous readiness for acquired responses, they are completely intelligible. In the first case I was in a state of readiness for locomotor responses adjusted to the actual configuration of the ground, but only that part of it to which the action was to be confined. My state of readiness, i.e. my consciousness, reflected true knowledge of that part of my environment which called for direct action but erroneous knowledge of the rest. In the second case my manual movements ranged over the whole table, and as it was in fact a plane surface, my movements had perforce to conform to that structure. The character of my newly acquired responses was reflected in my consciousness. I saw the table as flat. But because, in this situation, I was taking no direct action relative to the precise positions of other objects in the room, beyond reach of my hands, they continued to be seen as distorted.

Since, then, the facts of consciousness are explainable within the conceptual framework of the natural sciences, I maintain that it is Professor Burt who, by postulating a 'mind' that cannot be described in natural scientific terms, has rejected William of Occam's advice to scientists and philosophers, not to multiply *entia praeter necessitatem*.

CONSCIOUSNESS AND SPACE PERCEPTION

A Reply to Mr. Taylor's Criticisms

By CYRIL BURT

The results which Mr. Taylor has obtained from his ingenious experiments on space perception are highly instructive. Yet I fail to see how they conflict with the theory that I myself put forward: indeed, as I shall try to show, they seem rather to confirm it. May I, however, deal first with the basic differences between us?

I. In a recent paper on 'The Concept of Consciousness' (*Brit. J. Psychol.*, LIII, pp. 229-242) I ventured to repeat my plea for resuming the systematic study of consciousness as part of the task of psychology; and with this Mr. Taylor apparently agrees. The type of study I advocated was not merely observational or experimental, but quantitative and statistical; and I argued that many of the errors of statistical psychologists in the past arose largely from their neglect of the conscious aspects of the processes they investigated and of the checks that could be carried out by using introspection to supplement their numerical data. Mr. Taylor, on the other hand, maintains that the introduction of statistical methods into psychology is a mere "cloak for intellectual sterility" (*Brit. J. Psychol.*, XLIX, p. 106). And in his latest paper he now carries his criticisms still further.

He begins with an attack on statements which he supposes I put forward in a recent contribution (this *Journal*, XIV, pp. 145f.). My contention, so he says, was "that, in order to account for the phenomena of consciousness, it is *necessary* [his *italics*] to postulate an immaterial mind that is independent of the body". But may I ask Mr. Taylor to look again at the article he quotes? There I expressly stated that I "used the term 'mind' as a convenient label to designate the hypothetical basis of consciousness, whatever it may be"; and I went on to insist that "we should refrain from deciding by definition whether it was or was not material, or spatially extended, or even identifiable with certain parts of the brain". "Any detailed description of the mind" (I said) "was an issue not to be settled at the outset, but by deductions step by step from the conclusions reached as our investigations advance."

At the same time it seems clear that, if we hold that the basis of consciousness is what would commonly be classed as a material substance, such as the brain or some portion of it, then, in addition to its material properties, that substance must also be endowed with certain other properties (namely, a capacity for conscious experiences or relations) which are not ordinarily regarded as material. Thus, if a behaviourist who had learnt from Dr. Wilder Penfield the necessary surgical skills were to open up my brain while I was still conscious, he would fail to discover among the observable structures or processes anything resembling the sensations, thoughts, or feelings that I actually experienced. Indeed, the quality of conscious experience is so utterly disparate from the qualities we denote by the word 'material' that another adjective seems needed, and the most appropriate is surely 'psychic'. Mr. Taylor quotes this adjective, and adds "i.e., immaterial". But the equation of the two epithets is his, and not mine. The fact that a thing has psychic properties does not prove it to be "immaterial"; and I should be the last to hold that such an identification was "necessary". It might or might not be immaterial: both suggestions are merely tentative hypotheses; and the best we can do is to weigh the probabilities for and against each alternative, and carry out further investigations which may tilt the balance of probabilities decisively in one direction or the other.

According to Mr. Taylor, however, "if we say that 'this something is psychic or immaterial', we are locking the laboratory door and throwing away the key: for who can search for something that has no material existence?" But why assume that the only form of existence is material? Mr. Taylor admits that he sometimes feels pain; so long as it lasts, the pain surely exists, yet it cannot possibly be said to possess a *material* existence,

like his desk or his bottle of ink. Even the most thoroughgoing materialist postulates existents which are not themselves material. Physicists talk of 'fields of force' (electrical, magnetic, or the like), and today would, almost without exception, deny what many scientists of the nineteenth century postulated, that these fields must have a *material basis*. And my own suggestion was that what I had called the mind might possibly be interpreted in terms of a 'field theory'. Thus it is Mr. Taylor who has locked himself inside his laboratory, and thrown away what may prove to be the only key. Without pausing to examine the evidence or the arguments adduced, he bids us discard at the very outset anything resembling what Professor Broad has called a 'psychic factor'; and the only reason he gives is that it is not included among "the concepts that form part of the currency of natural scientific discourse". *A priori* arguments of this kind are by no means new. They were used by the astronomers who refused to look through Galileo's telescope; they were used by the physicists who refused to consider Einstein's theory of relativity or Planck's notion of discrete quanta; and they have formed the staple type of defence adopted by every behaviourist from Watson to the present day. To accept them as valid would be to block all scientific progress.

Unlike Watson, however, Mr. Taylor does not forbid the use of terms like 'consciousness' or 'awareness'. But he insists that I am wrong to think of consciousness as a 'simple and unanalysable relation'. Consciousness, he believes, "can be fully accounted for in behavioural terms". Let us therefore examine his reasoning. He takes a special case—perceiving the position of an ink bottle lying on the observer's desk. The physical stimulus, he says, "not only produces two distinct images on the two retinae of my eyes; it excites, or tends to excite, a behavioural response namely, stretching out my hand to grasp the bottle". These effects supply the necessary data; and he then undertakes a mathematical analysis of the problem of fixing the position of the ink bottle by means of its geometrical coordinates, referred to a "reference frame anchored in my body". He finds that a complete solution requires "an 8-dimensional subspace of the 11-dimensional apparent space". The measurements are obtained, not by using a tape-measure, but as a result of previously conditioned tendencies, which carry out the requisite computations (or their "analogues") in terms of "neurological findings". "The complicated mass of data has to be processed without benefit of the hardware built into a material computer. If there is a mind responsible for all this, the task would be not merely formidable, but forbidding... Professor Burt," he adds, "can afford to ignore all this, or make logical sport of it". But I neither ignored them nor made sport of it. Indeed, in discussing the process which mediate spatial perception, I myself emphasized the complexity of the underlying neural activities: my own example was that of a tennis-player receiving and placing a ball, and my comparison was with the computations of matrix algebra carried out by an electronic computer (*Psychometrika*, III, pp. 152f.).

So far we have encountered nothing that could be regarded as an actual perception or awareness. Mr. Taylor, however, continues his analysis by suggesting that, owing to the stimulation of these "conditioned reflex arcs or engrams", the organism is thrown into "a state of simultaneous readiness for a multiplicity of learned responses". And "finally", he says, "I postulate that this state of simultaneous readiness is identical with visual perception... I do not," he points out, "say that consciousness and simultaneous readiness are different *aspects* of the same thing: I say that they *are* one and the same thing".

But does this really demonstrate what he sets out to prove? The initial issue is somewhat confused because Mr. Taylor takes as an illustrative case an exceedingly complicated mental process, instead of the simplest conceivable instance, as I endeavoured to do. I chose the simple experience of pain. Let us apply Mr. Taylor's reasoning to a patient waiting to have a tooth extracted. Were Mr. Taylor the dentist, he would presumably assure him that there was no need for an anaesthetic. "The consciousness of pain," he would say, "is not, as you suppose, an irreducible experience. It is not a feeling, because feelings have no material existence. Pain is nothing but a state of readiness to make certain

behavioural responses—jumping up from the chair, crying out, and pushing the forceps away. And, since as a brave man you will have conditioned yourself *not* to make such responses, there can be no consciousness of pain."

This of course is a *reductio ad absurdum*; and it seems clear that there must be a fallacy somewhere. Nor is it far to seek. What Mr. Taylor originally undertook to do was to "show that awareness . . . can be fully accounted for in behavioural terms". At first this sounds as though he meant "in terms of *actual* behaviour"; it turns out however that his explanation is to be in terms merely of conditioned *tendencies* to behaviour; and these he supposes to exist in the form of material 'engrams' in the brain. Certainly, he says, Lashley failed to find any evidence for such 'engrams'; but the failure resulted from mistaken methods. Mr. Taylor makes no claim to have succeeded where Lashley failed, consequently, the 'engrams' and their 'activation' remain a sheer hypothesis. And it is with a hypothetical state of 'activation' in these hypothetical structures that Mr. Taylor proposes to identify consciousness. Why? Simply because he "postulates" their identity. Thus in the end, so far from "showing" that consciousness can be wholly reduced to overt behaviour or behavioural tendencies, Mr. Taylor is driven to "postulate" the very thing he set out to prove.

What then is Mr. Taylor's reason for asking us to accept such a paradoxical 'postulate'? It is because the assumption that consciousness is something ultimate and irreducible would introduce into psychology a concept which finds no place in "the currency of natural scientific discourse". But surely what concepts form part of a particular science is a matter to be decided according to the subject matter of that science. A physicist or a geologist has no need to include in his list of concepts the feeling of pain; but a chemist seeking to discover a new analgesic does need to include such a concept. He would not be content with "states of simultaneous readiness" as a substitute; yet that does not place him "outside the pale of genuine scientific investigators".

One or two of Mr. Taylor's supporters, however, have tried to suggest some more definite criterion. Mr. Rackham, for instance, has put forward the familiar distinction between what is 'public' and what is 'private'. Brains, engrams, states of readiness, overt actions—in a word, "material structures and their movements, together with the space and time in which they move"—these, we are told, are "things which in principle everyone can observe; they can therefore be verified; and they are in this sense 'public'. But a feeling of pain is a thing which only one man can observe, to wit, the person who feels it; it is therefore 'private', and consequently cannot be verified. We do not deny its occurrence: we merely say that, being unverifiable, it cannot be a subject for scientific study. Only by discarding such concepts can we succeed in abolishing subjectivity from psychology, and render it wholly objective, like the other sciences".

Nevertheless, popular as it is, such an argument will not hold water. In the first place, *every* observation, whether made by an introspective psychologist, a physicist, or Mr. Taylor himself, is 'private'. Yet that does not prevent us from verifying them. After all, what is it that we need to verify in the case of an introspection? Not the particular fact that, during dental operations, I or it may be John Jones felt a pain which could be obviated by an injection of procaine, but the more general fact that practically all normal individuals experience a similar pain and a similar relief. And since it has proved possible to establish a large number of generalizations about a wide variety of conscious experiences, we require a scientifically acceptable terminology to describe them. The generalizations thus demonstrated—e.g. the 'laws' of colour-mixture and colour contrast, the phenomena of after-images, reproductive imagery, and the like—are surely best formulated in terms of the sensory qualities with which we are all familiar. To translate them into statements about hypothetical 'engrams', or 'simultaneous multiple readiness' would be in no way helpful. Secondly, the view that "the currency of natural scientific discourse" excludes any reference to mental processes and mental contents is plainly based on what was "current" in nineteenth century science. This seems to be the notion of 'natural science' still retained by

most non-scientific writers; but it is not the notion of natural science offered to us by scientists of the present day. Nineteenth century science, though explicitly founded on observation, eliminated all mention of the observer. But the quantum physicists and the relativists have shown that, except where merely approximate conclusions are required, the observer can no longer be ignored. One example must suffice. Measurements of time play an essential part in all the physical sciences; and the physicist has learnt by bitter experience that the time which is measured is always "some observer's *private* time". It is true that, so long as we are concerned (like Mr. Taylor's Newtonian scientist) solely with man-sized objects and man-like speeds, all private times appear to agree. For example, in 1572 there was the startling outburst of the great Nova in Perseus reported by Tycho Brahe and verified by observers in London; in 1666 the Great Fire broke out: accordingly had Old Parr lived a little longer he would have reported that the Fire occurred 94 years after he had seen the Nova. But a space-traveller rushing along at a high but constant speed would have assigned a very different value to the interval; and an observer is conceivable for whom the two events would have happened simultaneously. Thus as Professor Bondi puts it, "each observer has his *own* private time; and, when high speeds are involved, it is the myriad private times that matter; nor do they combine to make a single independent public time". This is but one of many familiar illustrations of the radical changes that have overtaken the "currency of fundamental concepts" in the realm of natural science during the past half century, which psychologists still ignore. While the behaviourist has been striving to "abolish all subjectivity from psychology, and so render it wholly objective, like the other sciences", the more advanced branches of physics have become more and more subjective.

I have now sought to establish three main points: (i) that for an impartial and comprehensive study of human experience we are bound to adopt certain concepts based on *private* observation (concepts such as consciousness and its specifically mental contents and processes), and that, contrary to the common opinion, there is nothing whatever in contemporary science which forbids us to do so; (ii) that, when we carry out systematic studies of these contents and processes, we find that various generalizations emerge, such as can be formulated in the terms of what are called scientific *laws*; and finally (iii) that these laws and generalizations themselves must be expressed in terms of *conscious* processes: they cannot be reduced to laws or generalizations about known *material* substances or processes. We are thus led to postulate some kind of entity—some special type of 'substance', 'energy', 'event', or 'field'—which has a capacity for consciousness, and may therefore conveniently be referred to as a 'mind'.

Whether such minds are, or could be, independent of the owners' brains is a semi-philosophical issue on which any decisive conclusion appears at present out of the question. In seeking possible evidence perhaps the most promising lines of inquiry available are those suggested by recent researches in the field of parapsychology. Meanwhile, the safest plan for the theorist is to preserve an open mind, or, if he feels reluctant to sit on the fence, to accept provisionally the opinions expressed by philosophers like Professor Broad and Professor Price, who are familiar with both sides of the case. They tend, on the whole, to favour some form of dualism. That too is the view of the few impartial psychologists who have studied what meagre evidence there is, such as Dr. Thouless and Dr. Beloff. However, to avoid further misunderstandings, let me repeat what I have already said: "the dualism I envisage would *not* be a dualism of the Cartesian type"—the only type, it would seem, with which Mr. Taylor and Mr. Rackham are acquainted. In any case, contrary to Mr. Taylor's assertion, this was by no means an *essential* element in the hypothesis I myself put forward in the paper he criticizes.

II. Now may I turn to the more specific problem of spatial perception, and deal with the particular questions which Mr. Taylor has put to me in his present 'Note'? He invites us to "consider a number of points, A, B, C, . . . , strung out along a horizontal line in both objective and perceived space, such that A is to the right of B, B to the right of C, and so on. We perceive these points" (he maintains) "as an ordered series". And

his question is twofold: how do I account for the right to left ordering in 'perceived space' (i) under normal condition, and (ii) under certain artificial conditions which he describes in detail?

As thus stated, the problem is one first clearly formulated by Lotze (*Medizinische Psychologie*, 1851, pp. 351f.); and Lotze's solution—that of *Lokalzeichen* ('local signs'), interpreted in terms of specific movements and supplemented by a physiological basis of 'neural arcs' as suggested by McDougall—is the solution accepted by Mr. Taylor. The novelty in his hypothesis lies chiefly in the wording: where McDougall spoke of 'the formation of associations' and 'acquired physiological dispositions or neurograms', Mr. Taylor talks of 'conditioned reflexes' and 'engrams'. However, McDougall's notion of what he and Mr. Taylor call 'neural arcs' was limited to the little that could be guessed from anatomical investigations with the ordinary optical microscope; and that would seem to be the principal source of Mr. Taylor's description of his hypothetical 'neural structure' ("a network of neurons spread over widely separated parts of the cortex"). However, recent studies with the electron microscope, and the records obtained with micro-electrodes, have greatly added to our knowledge; and I should prefer to base my answer to Mr. Taylor's question primarily on the information so obtained. Since much of it seems unfamiliar to Mr. Taylor and his various supporters, and indeed is largely the result of researches published¹ since his book was written, perhaps I may be allowed very briefly to summarize it here.

The retino-cortical pathway may be regarded as consisting of six stages and including six types of neuron. Of these (1) three are in the retina, (2) one in the lateral geniculate body (at the back of the optic thalamus), and (3) two in the cortex.

(1) The primary neurons are the photo-sensitive elements in the retina—the rods and cones; the secondary, the bipolar cells of the inner nuclear layer, and the tertiary the ganglion cells whose fibres form the optic nerve. During a series of investigations started about a dozen years ago W. S. Kuffler of John Hopkins Hospital was able to demonstrate that the analysis of visual patterns begins not in the cerebral cortex, but in the bipolar and ganglion cells of the retina.² In the fovea the bipolar and ganglion cells are mainly of the 'midget' type (no 'mops' and no 'diffuse' or horizontal transverse cells): so that there is, as it were, almost a set of private lines. Most investigators, however, now seem agreed that the limits of visual discrimination are determined, not in accordance with the old hypothesis of geometrical correspondence of the images, but rather by the gradients in the diffraction pattern. Kuffler, however, succeeded in showing that there are two functional types of ganglion cell. Ordinarily the ganglion cells discharge at a fairly constant rate, even when there is no external stimulus; but in one type, owing to a slight but systematic overlapping with the photo-receptors, the stimulating field consists (a) of a small circular area in which an increase of light increases the rate of firing, and (b) a surrounding concentric zone where an increase of light diminishes the rate of firing; while in the other type of cell the effects of such stimulation are reversed. Thus here, in the retina itself, are the first steps towards perceptual localization.

(2) At the next relay station we find that many nerve fibres from the optic tract may funnel into a single nerve cell in the geniculate body. The responses of the geniculate cells show many of the characteristics of the retinal ganglion cells, but also recognizable differences. Their main function is apparently to increase still further the contrast between the responses to a small central point or spot of light and the responses to more diffused illumination.

(3) The fibres from the geniculate cells pass through the optic radiation to the so-called striped or striate area of the cortex which lines the lips and walls of the calcarine fissure

¹ See more especially D. H. Hubel, 'Integrative Processes in the Central Visual Pathways of the Cat', *J. Optic. Soc. Amer.*, LIII, 1963, pp. 58–66 and refs., and for earlier work R. Granit, 'The Visual Pathway', ap. H. Davson (ed.) *The Eye*, vol. II, Academic Press, 1962.

² 'The Functional Organization of the Mammalian Retina', *J. Neurophysiol.*, XVI, pp. 37–68

at the back of the brain. In the main they connect with the small cells of the thickened and highly specialized fourth layer (the internal granular layer, which is responsible for the 'stripe'), and so indirectly with the third and fifth layers, which in turn send fibres to neighbouring areas of the cortex and out of the cortex altogether.¹ From the work of Brouwer, Zeeman, Holmes and others, who have attempted to map the projection of the retina on the visual cortex, we know that there is a rough topological correspondence between minute areas of the two; but the correspondence is by no means one-to-one. Quite recently, however, D. H. Hubel and T. N. Wiesel,² working in the Harvard Medical School, have now carried the functional study of the visual pathways as far as the fifth and sixth stages. Their more important investigations were directed towards discovering how the information which the visual cortex sends out differs from the information it receives. The animal which serves as subject, usually a cat, is anaesthetized, so that all consciousness is presumably abolished, and placed before a wide screen about a yard and a half away from its eyes. One or more spots or patterns of light, sometimes stationary, sometimes moving, are then projected on to the screen; and by means of microelectrodes the impulses coming into and out of particular cells are recorded and compared. Here as elsewhere it is found that the processes at the synapses may be not only excitatory (as the theories of McDougall and Taylor tacitly assume), but also inhibitory. In the cortex, however, there are apparently no cells with concentric fields of reception, such as are found in the retina and geniculate body. Broadly speaking, it would seem that according to their functions the cells of the visual cortex can be divided into two principal types. What Hubel and Wiesel designate as 'simple cells' respond to linear stimuli—bright lines on a dark background, dark lines on a bright, and straight edges or boundaries between light and dark: but they only do so, if the line or edge is suitably placed and orientated for the specific cell. Thus, although several fibres from different cells in the geniculate body will stimulate one and the same simple cortical cell, to be effective the impulses from the geniculate cells must relate to light-stimuli in the same straight line. It is these 'simple' cells that chiefly make up the fifth stage in the total pathway.

What the investigators term 'complex' cells form the chief, but by no means the sole, elements in the sixth stage. These also respond best to linear stimulation (including contrast and contour), but are not so discriminative in regard to position; moreover, unlike a simple cell, a complex cell may be activated by moving lines, provided the lines do not rotate and pass out of the area connected with the cell in question: in short, *this type of cell continues to respond, even when the stimulus (the line or edge) moves across the external visual field, or, what amounts to the same thing, when the eye moves while the stimulus remains stationary.* One kind of complex cell has a suggestive property which deserves special mention: it always responds to a vertical edge, but shows an excitatory or an inhibitory response *according as the light area is to the right or to the left of the edge.* And here, so far as the neurological basis is concerned, we seem to have an adequate answer to Mr. Taylor's first question, namely that which relates to 'the perceived left-right ordering' of a set of visual stimuli.

From a functional standpoint it would appear that the visual cortex is divided into tiny columns, roughly cylindrical, and about 0.5 mm. across. In each column—in a sharply localized and restricted space—are gathered together all the cells whose fields have the same general position on the retina and the same orientation; and the functional connections between the several cells are, as we might have expected from histological studies, preponderantly vertical. Mr. Taylor warns the physiologist against "searching in a restricted locality for a network which is spread over a wide network". This may be sound advice if we are concerned with the total behavioural response; but clearly it is liable to be

¹Cf. D. A. Sholl, *The Organization of the Cerebral Cortex*, 1956, and refs.

²'Receptive Fields, Binocular Interaction, and Functional Architecture in the Cats' Visual Cortex', *J. Physiol.*, CLX, 1962, pp. 106-154.

misleading when we are discussing the simpler problem of localization within a purely visual field.

It will be noted that two characteristics I have emphasized—those which underlie the relative position and serial ordering of stimuli within the visual field, and those which underlie the identification of a stimulus as it moves across the visual field and the recognition of the stimulus as constant when the eyes move—are the result, *not* of prior conditioning, but of the arrangement of the neural structures in the retina-cortical pathway. As the investigators themselves point out, such characteristics are a natural consequence of the way the branching fibres of cells that belong to an earlier stage are connected with the cells of the next stage. The effects of stimulation, they say, can therefore be interpreted “in terms of quite simple concepts—nerve impulse, convergence of nerve-fibres, and their excitatory or inhibitory action”: and to understand such effects “requires no appeal to the behaviour of complex networks or of electronic computers, or to theories based on higher mathematics”. If that be so, then Mr. Taylor’s elaborate explanation in terms of “an 8-dimensional subspace of an 11-dimensional afferent space, further subdivided into a number of 5-dimensional subspaces”, and his insistence on the supplementary perception of “the directions of the eyes”, “the orientation of the head”, yielding “a complex pattern of events *conditioned* to the motor response of stretching out the hand to grasp the bottle”—all this becomes quite superfluous. Mr. Taylor contends that, unless these further data—“proprioceptive (i.e. kinaesthetic) rather than visual”—are also “registered in the afferent system . . . the position of the ink bottle cannot be perceived as constant”. But that is now seen to be incorrect.

May I offer my own interpretation of the processes involved? I shall assume that his ink bottle presents a blackish rectangular shape with vertical sides, contrasting with a relatively pale background, and that the spectator’s eyes are focused on a point in the light background to the left of the square. Then the left hand edge of the inkpot will, I imagine, activate those cortical cells which are preferentially excited by vertical contours with light areas to the left and dark to the right. If the spectator is interested in the location of the inkpot (as in seeking to dip his pen into it), there will no doubt be a reflex tendency to focus the eye on the inkpot itself. If his glance moves across the field, or if the inkpot moves while the eyes are stationary, some of the cells thus activated (including all the simple cells) will change; but a specific group of complex cells *will not change*, unless the vertical boundaries are sloped or move beyond the cells’ receptive field. Thus movement, or ‘readiness’ to make appropriate movements, are by no means necessary for localization within the visual field. Provided it was equipped with a nervous system similar to that of the higher mammals, Condillac’s sentient statue—which possessed only a single sense, namely vision, but no power of moving eyes, head, arms, hands, or legs—or (if such a fancy seems too far-fetched) a child completely paralysed from birth—would still be able to identify and localize the inkpot in a purely visual field.

Since Mr. Taylor is convinced that consciousness is impossible in the absence of an ability to make appropriate behavioural responses, he would of course reject such an account. But it now seems clear that, once again, he has really assumed in advance the very point he is professing to demonstrate. By insisting from the very outset on behaviouristic interpretations he has been led to confuse the problem of mere *visual localization* with the more complicated problem of *visuo-motor coordination*. In his analysis he therefore includes such “motor responses” as successfully “stretching out the hand to grasp the bottle”. Now an infant less than one month old can successfully follow a moving light with its eyes; but not until it is 4 or 5 months old will it succeed in grasping a ring that is dangled within reach of its hand. Here its eventual success no doubt must depend partly on maturation of the neural mechanisms for eye-and-hand coordination, and partly on practice and conditioning of the kind Mr. Taylor describes. But in such a case, as Piaget has pointed out in his studies of children’s perception of space, the ‘space’ with which we are dealing is no longer a purely sensory extension of the visual field, but a highly organized visuo-

kinaesthetic-motor space; and this is largely a conceptual space, and by no means merely perceptual. Hence the discussion of its detailed nature and development is really irrelevant to the simpler question in which we are here concerned.

Mr. Taylor, however, challenges me to explain one other paradoxical result which he thinks is crucial for his contention. When, with the aid of special optical apparatus, the 'structure' of the subject's visual field has been reversed, an adult subject can, after a period of re-conditioning, make the right motor responses, but "printed matter continues to be seen in reverse" i.e. he is unable to read the mirrored print. Mr. Taylor describes this by saying that "one portion of the field has reverted to its original structure, while the rest retains the new structure"; and he asks me "what sort of a field is capable of changing its spots in this odd fashion?"

The problem is a far more complicated one than Mr. Taylor's account implies. I have discussed the factors involved in several previous papers dealing (a) with the study of mirror-writing and mirror-reading in left-handed, right-handed, and so-called 'wordblind' children, and (b) with the problem of transfer of training after practise in drawing and writing, when the hand and the pattern to be made or copied are seen only in a mirror: (cf. *The Backward Child*, pp. 341-359 and refs.). To begin with, the subjects themselves, whether young or old, do not report any startling inconsistency in the 'structure' of their visual fields, such as Mr. Taylor's description would imply. The paradox only arises, because he insists on "considering the problem as a problem in coordinate geometry". Every object in the "field", whether inkbottle or printed letter, thus becomes a "pattern of ordered points"; and each point is supposed to have its own "parameters" or coordinates, expressing in "numerical values" or their "analogues" its position in an abstract "frame of reference". But that is not the way we build up our visual percepts or recognize objects and determine their location.

As Koffka has shown in experiments on space perception among children, the geometrical analysis of patterns has little to do with the accuracy of their recognition: "what counts is rather the meaning attached to the various complex shapes". Show a young child a series of drawings and then ask him to pick out those he has seen; and he will succeed best, not with simple geometrical shapes like a circle or triangle, but with more complex patterns to which he has attributed a meaning—e.g., a crude scheme suggesting a smiling face, or a cubical solid-looking box. Oddly enough, at a later age, when a more detailed analysis has become habitual, the child may fail where previously he succeeded. Thus, for most children who have learnt to read fluently, a printed word is not "a system of ordered points along a series of parallel axes" (as it is of course for the television screen): it is a distinctive shape with a meaning of its own—an idea, an object, or it may be just an articulate sound. The unit is not (as Mr. Taylor assumes) the constituent point, but the total word—or occasionally the syllable or the phrase; and there is only one relevant direction for the eyes to move—from left to right. Thus 'cat' is seen as a kind of ideogram signifying a well known animal; 'do' is similarly a shape signifying the sound *do*; viewed in the mirror, the reversed image of the former appears a meaningless muddle of strokes and curves, while the reflection of the latter is not recognized as 'do' printed backwards, but is seen as an entirely different shape signifying the sound 'ob'. Generally speaking, therefore, to read mirrored print we should have to start *de novo* and learn a different set of patterns, just as if we were learning to read Arabic or Hebrew. This mode of perceiving words, so far as my experience goes, holds good of practically all educated adults, like those who have taken part in the experiments of Kohler (referred to by Taylor) or in similar experiments in my own laboratory.

But in certain cases the unit is different. Thus children who have been taught by the old alphabetic method find it much easier to read mirrored words. I myself find mirrored handwriting quite illegible; mirrored numbers, as in dates, I can read with ease. I have little difficulty in deciphering Greek inscriptions written *boustrophedon*. And I can read Egyptian hieroglyphics almost as easily when they have to be read from right to left or from

above downwards, as when they are printed (as they usually are in modern transcriptions) from left to right; but this is because the hieroglyphics consist of phonograms or pictograms which have the same meaning whether or not they are reversed: the schematic hand which represents *d*, for instance, is just as easy to recognize which ever way the fingers point. Consequently, Mr. Taylor's generalization is much too sweeping; and on my theory the results he reports would present no paradox. They are indeed precisely what I should predict.

Moreover, his contention that sensory localization is only possible as a result of practice in motor localization breaks down completely when we turn from vision to other senses. I can locate a touch on the small of my back or a pain in my appendix, although they are in places that my hand could not possibly reach; and every student of psychology is familiar with experiments in which a blindfolded subject is asked to indicate with his finger the place where he hears a click, and then more often or not points to a position quite different from that of the stimulus. The sole effect of practice on sensory localization, when it has any, is to make it more precise and integrate one form of localization with others, not to make it possible.

In some early experiments I carried out with Sir Charles Sherrington at his suggestion, first the articular surfaces of my eyeballs, and then their kinaesthetic sensitivity, were anaesthetized with injections of cocaine; in these conditions I attempted to localize with a pointer spots of light which appeared in front of me in a dark room. Anaesthetizing one or other of the cues appreciably impaired the accuracy of motor responses: when both cues were anaesthetized, my attempts became astonishingly erratic. But the visual localization of the spot within the visual field was in no way altered.

However, even those who originally advanced a theory of associative local signs similar to that which Mr. Taylor now supports did not suppose that a knowledge of the underlying neural processes would supply a *complete* interpretation of spatial consciousness. Lotze drew the opposite conclusion; and in this he was followed by McDougall. "The activity", says McDougall, "call it creative or not as you please—which constitutes the final *knowing* of the position of a point, a star in the sky, for example—*can only be attributed to the mind itself.*"

But before we can determine the detailed nature of these final stages—those which link the actual 'knowing' to the related neural processes—far more research is wanted. And here, up to a point, I should agree with Mr. Taylor's mode of approach. It is a sound methodological principle to explain as much as we possibly can in terms of neurological activities before we go on to formulate supplementary hypotheses about the working of the mind. And, as Dr. Hubel, in the paper I have already quoted, emphasizes, the *immediate* need is to pass beyond the sixth stage to next stages within the brain, and observe how far that will help to elucidate the way we perceive and localize objects in visual space. Nevertheless, a complete interpretation solely or predominantly in terms of nerve-cells, nerve-impulses, and neural connections is, he admits, by no means easy to envisage. As he says, "one serious difficulty in such a scheme is that it would presuppose an enormous degree of cortical organization". Let us, however, suppose that such a scheme has in the end been satisfactorily worked out; there must always remain the last stage of all—the stage in which we pass from discrete currents in separate nerve-cells and separate nerve-fibres to the continuous visual field as we perceive it, with all its meaningful contents—a sensory field which itself is merged with the general 'stream of consciousness' that makes up the life-history of the experiencing subject. To compare the two in any detail, we shall have to wait until a second Wilder Penfield has made recordings with microelectrodes, placed not on the cortex of an anaesthetized cat, but on the cortex of a conscious human being. Yet, even when this has been achieved, it seems pretty obvious that the correspondence between the cortical processes and the conscious experiences will not be—in fact cannot be—exact, and that consequently part of the final organization will be the work of what I have called a mind or neuropsychic field. The assumption of a mind and of distinctive mental processes thus seems in the end inescapable.

SECONDARY SCHOOL SELECTION

To the Editor, *The British Journal of Statistical Psychology*

Sir,—May I send you some belated comments on the book 'Secondary School Selection' issued by a working party of the British Psychological Society under the editorship of Professor P. E. Vernon and published in 1957. I think it well that some misconceptions contained in the book should be removed even at this late date, for readers may have accepted all its statements and conclusions without question, as it derives from a committee of undoubted distinction in the field of psychology. I must add that the delay in making these comments comes from circumstances beyond my control.

I have always thought that the individual intelligence test, such as the Terman-Merrill, has acquired an undeserved esteem amongst psychologists as a measure of intelligence, whatever its merits for clinical purposes. One reads in the book that the Terman-Merrill test gives "an accurate assessment of an individual child", has "greater trustworthiness" (p. 94), and provides "a more reliable and representative all-round survey of a child's intellectual powers" (p. 95). Strangely enough, however, the authors continue, "there is no evidence that it is more predictive of ability to do grammar school work than is the result of a group test" (p. 95). Is there indeed any reliable evidence at all of its predictive value? The test is used in difficult borderline cases for want of better, but since its standardization relative to a particular education authority, both as regards mean and standard deviation, must be in doubt, and also since its correlation with a group verbal intelligence test having a validity of 0·8 is only about 0·83¹, not much reliance can be placed on its results. Incidentally the distressing wastage of information acquired in the course of the test offends one's parsimonious instincts.

The authors are not happy in their discussion of a criterion of grammar school success. They desire one containing estimates of personal development as well as intellectual attainments (p. 73), for "a child may profit more from a secondary school than his success in school examinations reveals" (p. 68). They overlook the fact that the criterion sought is one that represents the *difference* between the effects of grammar and modern school education, and if the two types of school develop character equally well, as is likely, the inclusion of personality traits in the criterion is vain. They require a multiple criterion and "not merely an averaged overall estimate" (p. 72), but make no suggestion as to how the elements of a composite criterion are to be used, if not pooled in some way or other. Their suggestion that more is wanted than a correlation between predictor and criterion "for the latter is a 'measure' of the power of the former" (p. 68) is not immediately comprehensible.

Here and there an unfortunate lack of precision is noticeable. For example, a statement such as "there is likely to be little difference in overall proportions of able children in Kent and in, say, Cornwall" (p. 17, footnote) is certainly unwarranted without sustaining evidence; the proportion of children of IQ not less than 115 may vary from about 10 per cent to 25 per cent according to district. Again, the tables on pages 81 and 82 cannot be interpreted, since the numbers from which the samples were drawn are not given. The recommendation that all children between the top 5 per cent and the bottom 50 per cent should form a border-zone (p. 177) is certainly questionable, since the position of the cuts will depend not only on the reliability of the examination but also on the level of ability in a district, and this can vary from a mean quotient of 94 to one of 105.²

Some of the tables in Appendix A might have been better chosen and most should have been more fully annotated. McClelland's table³ on page 178 calls for special comment,

¹ *The Trend of Scottish Intelligence*. Univ. of London Press, 1949, p. 123.

² See also A. E. G. Pilliner, "The Position and Size of the Border-line Group in an Examination", *Brit. J. educ. Psychol.*, 1950, 20, 133.

³ W. McClelland, *Selection for Secondary Education*. Univ. of London Press, 1942. Table XXXII, p. 71.

as it is and always has been misleading. The correlations were obtained from the average of the three correlations between (a) the predicting variables and (b) secondary school performance after 1, 2 and 3 years respectively. In my view performance at the latest stage, i.e. after 3 years is the only relevant criterion. Correction for selection was applied only to the second and third year results to bring them into line with those of the first year; the data therefore are not comparable with other correlations corrected to values applicable to the whole population. Finally, the intelligence test is the only non-composite variable in the whole table, and its appearance at the bottom of the order of merit might well be interpreted to its detriment by the unobservant. The last column of McClelland's Table XXIV (*ibid.* p. 67) is his only table relevant to the purposes of our book. Here the criterion is third-year performance, correlations of which with the several single variables are given. The intelligence test is now shown as the most predictive of all the single variables, although it occupied only 45 minutes against 125 minutes for the traditional English paper and 70 minutes for the arithmetic. McClelland gives no battery correlations for the third year criterion.

The two tables on page 180 extracted from Emmett and Wilmot's article¹ are not comparable, as the first table relates to correlations uncorrected for selection and the second to corrected data, the latter point not being mentioned. The substitution of the first table by the corrected correlations would have been more instructive. The table on page 181 should intimate that the criterion is based on secondary school teachers' estimates after 3 years in the grammar school and that the correlations are corrected for selection.

Richardson's tables on page 182 should have been recorded as corrected for selection. On page 134 we find mention of her comparison of 'home-made' English and arithmetic papers with Moray House tests. Reference to the article in question² shows that the validity of the former is judged against a criterion assessed after one year in grammar school (not "one and two years" as stated in the book) while the Moray House tests were correlated with criteria after five and three years respectively—an unfortunate comparison in the absence of fuller information.

Appendix B is a poor effort, obviously written by a lay statistician. Most of us educationists with statistical leanings should recognize that we are but amateurs at the statistical game, and should submit our writings to the expert scrutiny of a specialist statistician before publication; this appendix would certainly not have passed muster. It is concerned with "The Constancy of the IQ", or as I should prefer to say, "The Variability of IQ Measurements". Now different intelligence tests measure somewhat different mental functions, so that to be precise one should specify what particular type of test is in mind when variability of IQ is talked about. This view the authors repudiate; they regard the use of similar tests on the various occasions of testing as distorting the results (p. 184). I just cannot agree.

The use of probable errors with a constantly appearing factor of 0·6745 is archaic, and the definition of a probable error as the "median variation" is bewildering. One finds clumsy expressions such as "maximum range of variation" when simply "range" is meant, and "median maximum variation" which may mean anything. Algebraic symbols are sometimes used without adequate definition, and a quantity is said to be "raised by" another quantity when the second is obviously a factor and not additive.

An unfortunate choice of material from early workers for the estimation of variability seems to have been made. Much of this work does not comply with the necessary conditions for accurate estimates and might have been excluded. The correlations quoted, many lying between 0·7 and 0·8, certainly give a wrong impression of the reliability of modern

¹ W. G. Emmett and F. S. Wilmot, *The Prediction of School Certificate Performance in Specific Subjects*, *Brit. J. educ. Psychol.*, 1952, 22, 52.

² S. C. Richardson, "Some Evidence relating to the Validity of Selection for Grammar Schools", *Brit. J. educ. Psychol.*, 1956, 26, 13.

tests of ability. In contrast, pairs of Moray House verbal intelligence tests suited to differing ages from 11+ upwards and administered to several hundred children gave IQ differences corresponding to correlations of 0·94 after 2 months, 0·92 after 2 years, and 0·89 after 5 years.

The appendix as a whole is so perplexing that further comment is not worth while. My statistical consultant found it unnecessarily involved and in places incomprehensible.

W. G. EMMETT

A REPLY TO MR. EMMETT'S LETTER

Sir,—Mr. Emmett's comments certainly deserve attention, even if the book is by now somewhat outdated. I am grateful for his clarification of McClelland's and other results, though I see no reason to agree that a 3-year follow-up criterion is necessarily better than a combined 1st to 3rd year criterion. Also it is noteworthy that Dr. D. McIntosh, in his general survey of McClelland's results in 1948, bases his final conclusions as to the best battery for predicting secondary school success on the same table that Mr. Emmett criticizes.

I would disagree that there is any inconsistency in admitting that group intelligence tests give as good predictions of secondary school examination marks as the Terman-Merrill scale, or better, while yet holding that the Terman-Merrill is a better measure of intelligence, particularly in assessing emotionally disturbed pupils whose reactions to the group-testing situation are in doubt. There can be no satisfactory external criterion of the validity of the Stanford-Binet type of scale; it can best be judged in terms of construct validity. Perhaps the most convincing contributory evidence is the long-term predictive value shown in Terman and Oden's follow-up studies of high I.Q. children to adulthood.

Mr. Emmett's major objections are to Appendix B on 'The Constancy of the I.Q.', and I daresay that it might have been wiser to obtain the cooperation of professional statistician in writing it. However the whole book was submitted by the Working Party to the Council of the Society, which included at least two accepted statisticians, and several modifications were made in the light of their comments before it was published. I could not agree more that different intelligence tests measure different functions, and that the precise instrument used should always be stated in talking about a child's I.Q. But I would strongly disagree that the type of reliability that statisticians might prefer, based on the correlations between the same test given twice or between parallel or closely similar tests, are relevant to the discussion. Different tests have to be used if a child is retested after a couple of years or more; and even if the Terman-Merrill is employed, he will be tested at a different point on the scale on the second occasion.

The main issue then is that some psychologists and many educationists talk, or at least used to talk, of the I.Q. as being constant for most practical purposes, while others point to such work as Honzik's or Dearborn and Rothney's to prove that it is very far from constant. And the correlations between the same or parallel instruments given a short time apart provide no basis for assessing this variability. The literature surveyed in the Appendix is not "unfortunate"; it is the most relevant to the everyday interpretation of a child's I.Q., and much of it is constantly cited by those who wish to argue that the I.Q. is or is not highly variable. The survey aims to show that the various findings can be largely reconciled if account is taken of the many conditions that affect correlations between successive testings. Perhaps Mr. Emmett is no more familiar with psychological and educational discussions of the constancy problem than the present writer is with the niceties of statistical terminology. As for such phrases as 'median maximum variation', perhaps he would like to suggest a better term to describe the difference between the lowest and highest I.Q.s obtained by a typical or median child who is retested on numerous occasions.

P. E. VERNON

IS INTELLIGENCE NORMALLY DISTRIBUTED?

To the Editor, *British Journal of Statistical Psychology*

Sir,—May I raise a query about Table II of Sir Cyril Burt's paper in the last issue of this *Journal* (XVI, p. 182)? The table estimates the proportion of children with I.Q.s of 160 or over at 445 per million, but gives no indication of the accuracy of this estimate. The estimate itself is reached by way of the parameters for the Pearson Type IV curve obtained in Table I; consequently one might approach the question of accuracy either directly, or indirectly by considering the variation in the estimates of the Type IV parameters.

For the direct approach we may note that Table I shows that the sample of 4,665 children actually contained one child with an I.Q. above 160. One in 4,665 is 216 per million, and a Poisson calculation gives 5 and 1,194 as the one in forty confidence limits. On the one hand, these limits include the Table II estimate of 445, which is gratifying; on the other hand, they are extremely wide. Does the other approach enable us to narrow them?

The other approach seems to me to be beset with formidable difficulties. In the first place, the method of moments does not give efficient estimates, except in the normal case. Secondly, for Type IV the higher moments are infinite. I think one might perhaps circumvent these difficulties, and get some sort of assessment of the variation in the estimates of the parameters; but I have not persevered far with this, because, as it seems to me, a third and more serious difficulty looms behind. Why should the search be limited to the Pearson Types? In this case Type IV gives satisfactory agreement with the observations within the range where the observations exist; but outside this range some doubt is thrown upon it by the fact that it gives a non-zero chance for any I.Q. we like to mention, while common sense tells us that there must be some limit, though we do not know where that limit is. It is true that these chances are exceedingly small, and might for most purposes be neglected; but it is by no means clear that this is the case when we are concerned with extreme values.

It would be interesting to try a Gumbel curve for the extreme value problem; but I have no time to do this at present. More data, however, are really needed, as Professor Burt says in his final paragraph. In collecting them it would surely be a good thing to concentrate attention on the extreme values. The distribution of low and moderate values has comparatively little relevance in estimating the proportions of high values, because many shapes of the tail are compatible with much the same shape in the middle of the distribution. It would not be hard to locate most of the cases of high values without a systematic survey of the whole range, because most of them must be conspicuous by any standard. Some, no doubt, would be missed. None the less I should have thought that this would be a much more helpful approach to a reasonably accurate estimate than spending the same amount of labour in covering the whole range in fewer places. Any estimate for the whole of England would in any case have to be based on at least a dozen places; otherwise there would not be enough degrees of freedom between primary sampling units. In my work for the Ministry of Education I have usually taken between 20 and 30 Local Education Authorities when these have been the first stage of a three-stage sample, with schools as the second and pupils as the third stage. And I may add that there is an example of a different kind of 'extreme value' problem at p. 280 in the Newsom Report.

G. F. PEAKER

Reply

I agree with nearly all that Mr. Peaker has said, and hope that both the surveys and the calculations that he mentions may be systematically tried out. So far as circumstances permitted, I myself made several attempts to determine the number of children with high values for the I.Q. (*loc. cit.*, p. 187); and intended these to give a rough indication of the accuracy of the estimate which he queries. The relative frequencies showed a wide variation

from one locality to another. Had it been possible to include rural areas from Scotland as well as England, the range of variation would have been wider still, and have come very close to the 'confidence limits' that Mr. Peaker has computed.

However, my chief reason for starting with samples of the *whole* school population was the desire to check a more general hypothesis. It seemed to me highly probable, in spite of Pearson's objections, that individual differences in mental ability were the outcome of both multifactorial and unifactorial types of inheritance (not that I believe there is any *sharp* distinction between the two), and that the contributory causes were not entirely independent. On this basis it appeared to follow algebraically that the resulting frequency-distribution would be one given by the formula for Pearson's Type IV curve, and that the values obtained for β_1 and β_2 might serve to indicate the relative importance of the 'unifactorial' components. There was a further advantage in keeping to the Pearson types: these were the types most frequently used for other biometric data. The fact that Type IV yields "a non-zero chance for any I.Q. we like to mention" is to my mind a merit rather than a defect. It would be quite impossible to specify any finite limits: whatever value we take for the upper limit, it is always conceivable that a child with yet a higher I.Q. might sooner or later be found. 'Common sense', I fancy, is sufficiently satisfied when we recognize that an individual with an 'infinite' amount of intelligence would only occur if we were able to sample a population of 'infinite' size.

CYRIL BURT

OBITUARY

PROFESSOR NORBERT WIENER

Professor Norbert Wiener, mathematician, cybernetician, linguist, teacher and author, died in Stockholm of a sudden heart attack on March 18, 1964. With his death there passes one of the great intellects of our time.

He was born on November 26, 1894 in Columbia, Missouri, the son of Leo Wiener, first Professor of Slavic languages and literature at Harvard University. He was educated at Ayer High School, Massachusetts; Tufts College, Boston, where he received his A.B. degree at the age of fourteen; and at Harvard and Cornell Universities, receiving his M.A., and Ph.D. from the former at the age of 19. He was then awarded travelling fellowships by Harvard which enabled him to study with Russell and Hardy at Cambridge, and with Hilbert at Göttingen. Between 1914 and 1919 he held a variety of posts—as docent lecturer in philosophy at Harvard, instructor in mathematics at the University of Maine, at the General Electric works, Lynn, Massachusetts, as a staff writer for Encyclopedia Americana, as a mathematician at the U.S. Army, Aberdeen Proving Ground, and finally as a journalist with the Boston Herald from 1918 to 1919.

He began his long association with the Massachusetts Institute of Technology in 1919 as an Instructor in Mathematics, later as Assistant, Associate and Full Professor in Mathematics, and finally as one of M.I.T.'s select Institute Professors.

Wiener travelled widely, and from 1931 onwards he was Visiting Professor or lecturer at various institutions, among them Cambridge University; Tsing Hua University, Peiping, China; Instituto Nacional de Cardiología, Mexico; University of Paris; Indian Statistical Institute, Calcutta; Institute for Theoretical Physics and Cybernetics, Naples; and latterly, the Central Institute for Brain Research, Amsterdam.

He received academic and scientific honours from various bodies, among them the 1933 Bôcher prize in mathematical analysis from the American Mathematical Society. A few months ago he was one of five men to whom President Johnson awarded the National Medal of Science. Wiener was delighted, and proudly showed the President's letter to his friends. It is gratifying that he lived to enjoy this tribute to his achievements.

Among his publications (over 200 in number) were many books: *The Fourier Integral and certain of its Applications* (1933); *Fourier Transforms in the Complex Domain* (with

R. E. A. C. Paley) (1934); *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (1949); *Cybernetics* (1948, 1961); *The Human Use of Human Beings* (1950); *Ex-prodigy* (1953); *I am a Mathematician* (1956); *Non-linear Problems in Random Theory* (1948); *The Tempter* (1959), and *God and Golem, Inc.* (1964).

Wiener was principally a mathematical analyst, although his earliest papers were concerned with algebra and logic. In all his later work he was strongly influenced by insights drawn from his observations of physical phenomena. In 1921 he published the first of his many papers on the mathematics of Brownian motion, after musing by the side of Boston's Charles River on the nature of the irregularities in its wavemotion. This initiated a series of papers on the theory of integration in function-spaces, in which certain Gaussian probability measures, now often called Wiener measures, were introduced to describe the Brownian motion. This work was closely linked with Wiener's studies of the Fourier Integral which he used in his development of Tauberian theorems, a proof of the Prime Number Theorem, a theory of general Harmonic Analysis, and in his later work on Ergodic Theory. At the same time, he initiated new developments in Potential Theory, which have since been incorporated into the modern Theory of Potential. More recent work by Doob, Fekman, and others, has vindicated Wiener's physical intuitions, and the Wiener measure and the Wiener integral have been used extensively in Newtonian Potential Theory and in Quantum Theory.

His interest in all things oscillatory led him to communication engineering, and it was natural that he should play a large part in the revolution which transformed this branch of engineering from an art into a discipline possessing a good measure of exact and powerful techniques. Wiener's work on wave-filters, on prediction, filtering and control apparatus for anti-aircraft systems (wherein an integral equation, now known as the Wiener-Hopf equation proved to be of fundamental importance), and on the general problems of feedback control mechanisms, dates from this period—the late 1930's and early 1940's. It is here that his earlier work on Harmonic analysis, initiated by physical considerations, and based solidly on the work of Fourier, Gibbs and Heaviside, came full-circle back to the design of physical hardware. This linear theory is now a basic item in the electrical engineer's repertoire. More recently, he used his theory of Brownian motion and stochastic integration to develop a non-linear theory, but there has been but little practical application of this to date.

Wiener was a man with a very wide circle of friends and colleagues, among them many scientists concerned with animal behaviour and neurology. In 1943, he collaborated with Julian Bigelow, an engineer, and Arturo Rosenbleuth, a neurophysiologist, in writing a short paper on the role of feedback in purposive behaviour, and so began the science of Cybernetics, i.e. the comparative study of control and communication in animals and machines. Together with Bigelow, he recognized that what is important in controlled behaviour is not so much the feedback of energy, but the feedback of specific signal patterns, or messages. He had also been studying the requirements for high speed computation by digital methods, and was led to a consideration of the analogies existent between the neural networks of the brain and those switching networks which were beginning to be used for high speed computation. Wiener presented his ideas on these subjects in the now famous book *Cybernetics*, published in 1948. A highly personal book, it has proved an immense stimulus for much brilliant interdisciplinary work—and alas, in the hands of men of less comprehensive intellect than Wiener's, has also stimulated work with no discipline at all. Herein was defined (independently of C. E. Shannon, in his theory of communication in the presence of noise) a measure of information closely related to J. Willard Gibb's measure of entropy, around which a theory of signal processing and transmission has since been constructed. With this work by Wiener and Shannon, scientists and engineers gained a sensitive and reasonably plausible (but by no means all-embracing) measure of what is relevant or irrelevant ("signal" or "noise") in situations wherein organisms, or machines, interact with their environment.

Since then, of course, Cybernetics and Information Theory and the companion subjects or Automata Theory and Systems Theory have developed and diversified in multifarious ways. If their lasting impact on the fields of Neurophysiology, Psychology, General Physiology, and the like, has not been as spectacular as was initially expected by the cyberneticians, it is fair to say that very many "orthodox" scientists have been stimulated by, and have used the notions, if not the theorems, of these communication sciences to great effect.

Wiener was quite concerned about the socio-economic aspects of the automation of industry, which he predicted would follow the development of these sciences, and made many efforts, seemingly not too well received, to alert a wider public to the need for long term planning of the next Industrial Revolution. Perhaps he lacked perspective on the reaction of individuals to automation, but his warnings are nonetheless important—even more so now.

There are many stories in the Wiener mythos, and we recount here only the one we think the best—and make no claims for its veracity. One day an M.I.T. professor saw a large number of the most brilliant mathematicians from M.I.T. and nearby Harvard milling around Wiener's office. He asked one of them what it was all about. "Wiener has verified the Riemann Hypothesis, and everybody's come to see his proof". Excitedly, the professor joined his colleagues to witness this turning point in mathematics. Wiener started to talk, chalking up an imposing array of integrals and Fourier transforms and infinite series on his blackboard. The room was tense with excitement. Then Wiener stopped, stood back, looked over all he had written, puffed intently on his cigar, stood awhile in thought, and then paced reflectively around the room. Suddenly he announced: "It's no good, I've proved too much—I've proved there are no prime numbers." !!

We, who as graduate students, had the privilege of meeting the great man when he still roamed the corridors of M.I.T. capturing stray listeners with a flow of fascinating ideas and anecdotes, will always recall with pleasure the image of this short, rather quaint figure, possessed of bifocals and a large cigar, walking splay-footed around the Institute, or regaling an audience during a conference recess with vaudeville songs of Boston's 1920's. He was called "America's oldest child prodigy", and if some used the description unkindly, to most of us this was an affectionate way of describing his few shortcomings which became so understandable when one read of his strange scholarly childhood in his autobiographical *Ex-Prodigy*. Even at the age of 69 he was still extremely active intellectually. Last February he was to be found in Amsterdam, learning Dutch, pouring forth ideas on Quantum Theory and the Brownian motion, Lasers in Biology, and Brain modelling, full of enthusiasm for the world within and without him. It was a great shock to hear of his death, when the memory of his vitality of a few weeks before was still fresh in our minds. It is hard to believe that he is dead, and that he will never again propound a new idea to a student, beam at him, and say, "Now isn't *that* interesting" and wait to be told that it is. It is a loss to us all that he sings no more.

J. D. COWAN
M. A. ARBIB

BOOK REVIEWS

Psychology through Experiment. Edited by GEORGE HUMPHREY. London: Methuen. 1963. Pp. 307. 30s.

I think this is a useful book although I doubt whether it accomplishes quite what its editor intended. It is said to have two aims. One is to provide "a companion volume for a practical class doing experiments . . . The second . . . is to induce those primarily interested in the humanities to take a glimpse at the kind of thing going on in practical classes". The contributors themselves seem to have been concerned to fulfil the first of these aims rather than the second. Thus technical terms are used without being defined and the presentation is often much too condensed for the reader with no background in psychology. Another barrier to the general reader is that there is no introductory section on experimental design or on the role of statistical techniques.

Regarded as a companion to a practical course the book is certainly more successful. Its weakness here lies in its heterogeneity. Hunter's chapter on learning is clearly intended for the beginner. (He is perhaps the only contributor who has tried to keep the general reader in mind). Yet Deutsch's chapter on experiments with animals is suitable only for the relatively advanced student. Another aspect of the book's heterogeneity is that the authors vary considerably in their willingness to use statistical techniques where the assumptions of the techniques may not hold. The introductions to the experiments are often extremely well written; and the book has considerable merit simply as a collection of essays on psychology. Indeed, Gregory's excellent chapter on sensory processes describes a number of ingenious demonstrations but no formal experiments. Other chapters are on perception (N. F. Dixon), motor performance (H. Kay and J. Szafran) and thinking (B. M. Foss).

Some individual points call for comment. Page 73 contains a very inadequate definition of information, "information is defined as the logarithm of the number of alternatives, the use of logarithms being convenient". On page 149 an incorrect method of correcting for guessing is given: if the ratio of old items to new items in a recognition list is 1 to 2, the expected ratio of lucky to unlucky guesses is 1 to 2 only if no old items are eliminated without guessing. On pp. 151-2 it is stated that two sets of scores can be compared by calculating "the probability that the obtained difference could have arisen merely by chance": the student cannot be assumed to understand that "the obtained difference" is short for "a difference as large as the difference obtained". The statement at the top of page 154 is dubious in view of the evidence that either recall or recognition can affect subsequent recall or recognition. Finally, on page 253 there is the arresting statement "It is possible that the motor theory of thinking holds only for certain extreme types of people". This suggests that some people think with their muscles and that others think with their nervous systems. However, perusal of the context shows that what is meant is simply that movements may accompany thought in some people only.

JOHN BROWN

Some Aspects of the Relative Efficiencies of Indian Languages. By B. S. RAMAKRISHNA, K. K. NAIR, V. N. CHIPLUNKAR, B. S. ATAL, V. RAMACHANDRIAN, R. SUBRAMANIAN. Bangalore: Department of Electrical Communication Engineering, Indian Institute of Science.

Indian, with its numerous independent languages spoken by large populations for whom English is a common secondary language, provides a natural laboratory for comparative investigations of the translation process. These opportunities have been exploited admirably in this little monograph, in which the authors have developed methods based on the information measure for evaluating quantitatively the relative efficiency of translation between English, as a common reference language, and six of the most important Indian languages (Hindi, Marathi, Tamil, Malayalam, Telugu, and Kannada). Their study

merits attention for its clear and intelligent development of the information measure as a method for evaluating translation, quite apart from its more topical implications regarding the relative advantages and disadvantages of the specific languages examined.

Their approach is to calculate the entropies of "semantically equivalent" texts in English and in each of the Indian languages. Recognizing that translation from one language to another inevitably introduces difficulties of expression, they give calculations both for translations into English and from English into each Indian language. From the difference between the entropies obtained in opposite directions of translation, they derive a measure of the facility with which semantic content can pass between any given pair of languages, a quantity that may well prove of considerable value in evaluating the similarity, structural and semantic, of different natural languages. Their results indicate that English encodes semantic content more efficiently than any of the Indian languages, of which Tamil is the most efficient and Hindi the least. From the point of view of telegraphy these differences can be considerable: nearly 25 per cent more time is required to transmit a given message in Hindi than in English.

A second issue the authors have attacked is the relative efficiency of writing Indian languages in their conventional scripts or in the Roman script adapted from English. This part of their work is probably of less general interest, and involves experimental procedures subject to more serious objections than the work on translation. But their method of attack is always sensible and often ingenious, as in their way of evaluating practice on writing speed in the different languages.

Throughout their work the authors are scrupulous in pointing out the limitations of their methods and modest in putting forth claims for their conclusions. The exposition of the logic behind their methods is exceptionally clear and concise, traits that are particularly welcome in a field where they are so rare.

D. H. HOWES

Bibliography of Basic Texts and Monographs on Statistical Methods. By W. R. BUCKLAND and R. FOX. Edinburgh and London: Oliver & Boyd, 1963. Pp. vii + 297. 35s.

In 1949 the International Statistical Institute introduced a programme entitled International Statistical Education. One part of the programme, which was sponsored and partly financed by UNESCO, was concerned with the preparation of various aids to teaching and research, including special bibliographies. The first of these bibliographies, originally published in 1951, now appears in this revised and enlarged second edition.

About 200 basic texts and monographs are listed, all in English. Most were written in the fifteen years up to 1960, but earlier works of greater importance are also included. Each entry gives the chapter headings of a book and then quotes extracts taken from reviews appearing in statistical and allied journals. There is also a supplementary list giving titles of works appearing in 1961 and 1962.

Twenty-two periodicals were consulted in compiling this work. As far as psychological journals are concerned the sample taken is neither suitable nor representative. Certainly British psychology does well, since the *British Journal of Psychology*, the *British Journal of Educational Psychology* and *Occupational Psychology* are all quoted, but this *Journal*, the technically relevant one, is not. No American psychological journals were considered, an especially remarkable omission being *Psychometrika*. As a consequence psychologists will not find much reference to texts which are concerned with their particular methodological problems, and several well-tried primers receive no mention. Therefore this work may not be a very satisfactory aid to the psychologist involved in undergraduate teaching. On the other hand, the research worker who is used to independently exploring the statistical literature will find that this bibliography provides an easy means of locating suitable expositions of a wide variety of topics in statistics and probability theory.

R. J. AUDLEY

Statistical Estimation in Factor Analysis: A new technique and its foundation.

By K. G. JÖRESKOG. 1963. Uppsala, Almquist and Wiksell. Pp. 145.

The subtitle of this monograph is an essential part of the title, as the author puts forward a new factor analysis model, and discusses in detail the problems of statistical estimation and hypothesis testing that arise. The book is thus essentially a research monograph, and requires careful consideration before it could be recommended to a wider audience than statisticians and psychologists actively concerned with the methodology of factor analysis.

Broadly speaking, the model assumed is still the orthodox linear multiple factor model, but a new restrictive assumption is introduced, *viz.* that the residual variances, after removing the effects of the group factors, are proportional to the reciprocal values of the diagonal elements of the inverted population dispersion matrix. This assumption, while it is claimed to lead to simplification in the inference problem (such as the avoidance of iterative techniques of the kind necessary in Lawley's maximum likelihood method), appears to be quite arbitrary, and to be advocated not so much for any intrinsic plausibility as for technical convenience. To take a simple case of one general factor and three tests, so that the (standardized) factor model is given, say, by

$$\begin{aligned}x_1 &= m_1 f + s_1 \sqrt{1 - m_1^2} \\x_2 &= m_2 f + s_2 \sqrt{1 - m_2^2} \\x_3 &= m_3 f + s_3 \sqrt{1 - m_3^2}\end{aligned}$$

we have the reciprocals of the diagonal elements of the inverted matrix as

$$\Delta/(1 - m_2^2 m_3^2), \quad \Delta/(1 - m_1^2 m_3^2), \quad \Delta/(1 - m_1^2 m_2^2) \quad (1)$$

where

$$\Delta = 1 - (m_1^2 m_2^2 + m_2^2 m_3^2 + m_1^2 m_3^2) + 2m_1^2 m_2^2 m_3^2,$$

compared with the true residual variances in this case

$$1 - m_1^2, \quad 1 - m_2^2, \quad 1 - m_3^2. \quad (2)$$

However, the technique appears to give a very good account of itself when tried on actual data, so that it can hardly be dismissed offhand. The data tested out are of two kinds, artificial and empirical. The use of artificial data is not of course to be deprecated, as it enables the analysis to be tried out under conditions which are precisely known. An extensive series of experiments was made, ranging from 10 to 30 tests and 2 to 6 group factors. It is important to notice that these artificial series were not necessarily restricted to satisfy the extra assumption, a restriction which would have been open to criticism. It is remarked in Chapter III that this implies a specification error in the residual variances, but it is claimed (p. 71) that this introduces a specification error in the factor loadings of a smaller order or magnitude. One 'ill-conditioned case' is discussed in this chapter, but this is a case of approximately equal latent roots leading to factor identification problems, which are largely resolved by rotation of the factors to a more stable position. What seems less conclusive is the effect of the specification error on the testing procedure discussed in Chapter IV. One would expect the goodness of fit test for the number of factors required to break down for large samples, but if this point was discussed the reviewer did not notice it.

In the final comparisons on empirical data, the most interesting comparison is with Lawley's maximum likelihood method, using two sets of data:

- (i) the data used by Emmett with 9 tests and three apparent factors (sample of 211 children);
- (ii) a battery of 33 tests used by Lord with at least ten apparent factors (sample of 649 students).

There is no doubt that the new method does extremely well on these two examples, so that as a 'working method' the value of the new technique seems established. How will it compare ultimately with Lawley's maximum likelihood method? The author himself says of this latter method (p. 129):

"This method is definitely the best available method for anyone who could afford to carry it through."

Thus the value of the new technique would seem at present to lie either as a useful self-contained procedure for those unable to embark on the full maximum-likelihood method, or perhaps as an initial computation that could be used as a convenient starting point for the more laborious method. Before it can be assessed as a foolproof final procedure in its own right, some further elucidation and discussion of the effects of the specification error resulting from the extra assumption would still appear to be needed. M. S. BARTLETT

Readings in General Psychology. Edited by P. HALMOS and A. ILIFFE. London: Routledge and Kegan Paul. 1963. Pp. x+251. 25s.

The editors point out that the student may be introduced to the study of human psychology in two ways—first, by an elementary but systematic study of the whole areas, secondly, by an intensive study of half a dozen topics only; in issuing the present collection of papers, they explain that their own aim has been to produce a textbook which follows the second line of approach. But would any teacher adopt such a procedure in the case of (say) human anatomy and physiology? One great merit which the editors claim is that “all fifteen papers were written by British authors”, so that the student is introduced to the subject “in a homelier idiom than it wears in most American textbooks”. The papers themselves however are of very unequal value. There are admirable discussions of cerebral localization by Lord Adrian, factor analysis and intelligence by Sir Cyril Burt, and the assessment of personality by Professor Vernon, an article by the late Dr. Ernest Jones on the ‘normal mind’ (published over 30 years ago), and some very indifferent contributions by persons who do not even seem to be members of the British Psychological Society. The book concludes with a paper on ‘psychology and ethics’ written especially for the volume by Mr. Halmos. He concludes that ethical statements, such as “this behaviour is right” can be resolved into a set of “psychological propositions about the nature of man”: for, “if a man refuses to guide his conduct by the demands of his essential nature, he would not survive; and the life sciences, including psychology will decide what is essential to survival.” This seems an extraordinary doctrine for psychologists to preach to their students. In a recent trial arising out of the crimes at Auschwitz the judge maintained that, if the only alternative to committing certain crimes, was to die, then in such cases a man ought rather to choose death. Mr. Halmos apparently would say that ethics and psychology would advise him to choose the crime and so survive.

J. P. MILLER

Readings in Psychology. Edited by J. COHEN. London: Allen & Unwin, 1964. Pp. 414. 52s. 6d.

This volume consists of twenty-four articles nearly all of which have been previously published, but have been revised for the present edition. The aim apparently has been to give students in this country a sample of the divergent views of the leading English psychologists actively writing at the present. Thus almost every English professor contributes a paper. The result is, as the editor observes, a wide diversity of outlook ranging from extreme materialism of the followers of Watson and Skinner to the neo-mentalism (if I may coin a phrase) of Professor Burt. The behaviourists rather spoil their case by their unattractive style; the papers by writers with a more philosophical approach, like Professor Mace and Professor Burt, are by far the most readable. But two most suggestive contributions of all are by contributors who are not psychologists—namely, Dr. B. V. Bowden who writes a most lively chapter on ‘Thought and Machine Processes’, and Professor Mackay who writes on ‘Information Theory and the Study of Man’—a paper which every student of statistical psychology should study. All the main aspects of psychology are represented; and the series, in spite of its miscellaneous character, is arranged in a logical and progressive order. Although the book is rather expensive it is admirably produced and printed.

J. P. MILLER

INDEX TO VOLUMES I TO XIV

An index to the first fourteen volumes of the *British Journal of Statistical Psychology*, classified according to authors and subjects, is now available. As proprietors of the journal the British Psychological Society wish to take this opportunity of expressing their gratitude to Mrs. Beatrice Warde to whose generosity and expert assistance the compilation and printing of the index is due.

OCCUPATIONAL PSYCHOLOGY

Editor : ALEC RODGER

CONTENTS

VOLUME 38, No. 1

JANUARY, 1964

- A Manpower Study : the Chemist in Research and Development. By DENIS PYM
Individual Differences in Overtime Working. By SYLVIA SHIMMIN and GWYNNE DE LA MARE
Training the Middle Aged for Inspection Work. By EUNICE BELBIN and SYLVIA SHIMMIN
An Autobiography. By WINIFRED RAPHAEL
Book Reviews.

Annual Subscription 40 shillings

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1

2 DEC 1964

BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

Vol. XVII Part 1.

May 1964

CONTENTS

	<i>Page</i>	
LUCE, R. DUNCAN	Asymptotic Learning in Psychophysical Theories	1
TREISMAN, MICHEL	The Effect of One Stimulus on the Threshold for Another: an Application of Signal Detectability Theory	15
KEATS, J. A.	A Method of Treating Individual Differences in Multi-dimensional Scaling	37
DAS, RHEA S.	Item Analysis by Probit and Fractile Graphical Methods	51
HENDRICKSON, ALAN E. AND WHITE, PAUL OWEN	Promax: a Quick Method for Rotation to Oblique Simple Structure	65
NOTES AND CORRESPONDENCE		
	What is Consciousness?	71
	Secondary School Selection	86
	Is Intelligence Normally Distributed?	89
Obituary	Professor Norbert Wiener	90
BOOK REVIEWS		93

Vol. XVII, Part 2

November 1964

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
R. J. AUDLEY

WITH THE ASSISTANCE OF
CYRIL BURT

AND THE FOLLOWING EDITORIAL BOARD

C. BANKS	A. R. JONCKHEERE
M. S. BARTLETT	D. N. LAWLEY
G. H. BOWER	R. D. LUCE
D. E. BROADBENT	S. MESSICK
J. BROWN	E. A. PEEL
V. R. CANE	L. S. PENROSE
D. R. COX	J. FRASER ROBERTS
L. J. CRONBACH	W. STEPHENSON
C. I. HOWARTH	A. STUART
P. E. VERNON	

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20s. per part (U.S.A. \$3.50)

Subscription 30s. 6d. per volume (U.S.A. \$4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The *British Journal of Statistical Psychology* is issued by the British Psychological Society. The subscription price per volume is 30s. net for Members of the Society and 30s. 6d. (post free) for non-members. The subscription price in the U.S.A. is \$4.50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1: non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, about May and November.

Papers for publication and books for review should be sent to the Editor, R. J. Audley, Department of Psychology, University College London, Gower Street, W.C.1. Authors should retain copies of their manuscripts as neither the Society nor its Editors can undertake any responsibility for loss or delay in the case of unsolicited contributions. Authors will receive 50 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology*, at a subscription rate of 60s. per annum for either journal, and *The British Journal of Social and Clinical Psychology* three times a year at a subscription rate of 50s. per annum. Subscriptions for these journals should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology* three times a year: for this the subscription, 30s. per annum, should be sent to Methuen & Co. Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1.

Papers for publication in *The British Journal of Psychology* should be sent to Professor Arthur Summerfield, Department of Psychology, Birkbeck College, Malet Street, London, W.C.1, those for publication in *The British Journal of Medical Psychology* to Dr. Thomas Freeman, Lansdowne Clinic, 4 Royal Crescent, Kelvingrove, Glasgow, C.3; those for publication in *The British Journal of Social and Clinical Psychology* to Mr. Michael Argyle (Social Psychology Editor), Institute of Experimental Psychology, 1 South Parks Road, Oxford, or to Dr. Ralph Hetherington, Department of Studies in Psychological Medicine, 6 Abercromby Square, Liverpool 7; and those for publication in *The British Journal of Educational Psychology* to Mr. L. B. Birch, University of Sheffield, Institute of Education, Sheffield 10.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer: that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract: (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the 'Recommendations to Authors' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21-73), or to those given in the paper on 'Setting Mathematics' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost. Contributors should note that the *Journal* is intended for the general psychological reader, not solely for those who specialize in statistics. Unusual technicalities should be either avoided or this *Journal*.

In future authors will receive one complete slip-proof of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible. All manuscripts should be submitted in duplicate.

FACTOR TRANSFORMATION METHODS

By D. N. LAWLEY

Department of Mathematics, University of Edinburgh

and

A. E. MAXWELL

Institute of Psychiatry, Maudsley Hospital, University of London

Methods are described for transforming factors in such a way that the final loadings are in reasonable conformity with a given *pattern* in which some of the elements are postulated to be zero. A distinction is made between factors which, in their transformed state, are (i) correlated and (ii) uncorrelated. The methods are illustrated by numerical examples.

I. PROBLEM

Much controversy amongst factor analysts has arisen because the matrix of loadings derived from the analysis of a given covariance or correlation matrix is not unique. The factors can be 'rotated' within the factor 'space', and may or may not be kept orthogonal (see [7], pp. 5–57 *et seq.*). One of the aims of rotation is to reduce to zero, or near zero, as many as possible of the loadings since this reduces the number of parameters required to describe the data and facilitates interpretation of the factors. Various empirical methods for achieving this end have been proposed of which the *varimax* method [5] is the best known. It achieves a simplified and unique orthogonal solution by rotating in such a way that, for each factor, differences between loadings are maximized. If the factors are allowed to become oblique, or correlated, then the *promax* method of rotation [3] may be employed. When these rotation procedures are used the hope is that the factors obtained will be readily interpretable. However, since the interpretation of factors depends ultimately on prior knowledge of the nature of the variates being analysed, it has also been proposed [2, 4, 6, 7] that the experimenter should be asked to postulate in advance, for a given set of data, a factor pattern in which the positions of the zero or near zero loadings are indicated. Given this information, efficient methods of estimating the non-zero loadings are available (see [7], chapter 6). In the estimation procedure, which is iterative, it is desirable to start with fairly good initial estimates of the loadings. Methods of obtaining these, for both correlated and uncorrelated factors, are given in [7], chapter 5; but as these methods are not very easy to programme for electronic computers, more suitable methods are considered here.

II. TRANSFORMATION TO CORRELATED FACTORS

We begin with a $p \times k$ loading matrix \mathbf{L} , of p variates x_i on k factors f_r . We wish to transform to new factors f_r^* such that the loading matrix \mathbf{L}^* on these has zero (or near zero) values for some of its elements. The positions of these zero values are determined by the experimenter who is asked to supply a $p \times k$ 'pattern' matrix $\mathbf{Q} = [q_{ir}]$ in which postulated zero loadings are represented by zeros and non-zero loadings by unities. The original factors f_r are assumed to be orthogonal or uncorrelated but, in the first instance, we shall suppose that the transformed factors may be correlated. The problem is to find a $k \times k$ transformation matrix \mathbf{M} such that $\mathbf{L}^* = \mathbf{LM}$ has zero or small elements in positions corresponding to zero elements of \mathbf{Q} . \mathbf{M} must be non-singular, and for \mathbf{L}^* and \mathbf{M} to be unique a necessary (though not sufficient) condition is that \mathbf{Q} should have at least $(k-1)$ zero elements in each column. An iterative solution to the problem is first given, but it is then shown how a non-iterative solution may be obtained.

To obtain a first approximation let us choose \mathbf{M} in such a way that each column of \mathbf{LM} is, when multiplied by a scaling factor, in maximum conformity with the corresponding column of \mathbf{Q} , in a Least Squares sense. This gives [1]

$$\mathbf{M} = (\mathbf{L}'\mathbf{L})^{-1}(\mathbf{L}'\mathbf{Q})\mathbf{D},$$

where \mathbf{D} is a diagonal matrix chosen so that

$$\mathbf{P} = (\mathbf{M}'\mathbf{M})^{-1}$$

has unit diagonal elements. We then have

$$\mathbf{LL}' = \mathbf{L}^*\mathbf{M}^{-1}\mathbf{M}'^{-1}\mathbf{L}^* = \mathbf{L}^*\mathbf{PL}^*.$$

Hence the correlation matrix for the transformed factors is \mathbf{P} .

The first approximation to \mathbf{M} can be improved by an iterative process. Let a suffix 1 in brackets attached to the matrices \mathbf{M} , \mathbf{L} , \mathbf{D} and \mathbf{P} in the preceding paragraph refer to the first iteration of the process. Let $\mathbf{L}^*_{(0(1))}$ be the matrix formed from $\mathbf{L}^*_{(1)}$ by replacing an element by zero whenever the corresponding element of \mathbf{Q} is zero. To obtain a better approximation for \mathbf{L}^* let $\mathbf{M}_{(2)}$ be chosen in such a way that each column of $\mathbf{L}^*_{(2)} = \mathbf{LM}_{(2)}$ is in maximum conformity with the corresponding column of $\mathbf{L}^*_{(0(1))}$. This gives

$$\mathbf{M}_{(2)} = (\mathbf{L}'\mathbf{L})^{-1}(\mathbf{L}'\mathbf{L}^*_{(0(1))})\mathbf{D}_{(2)},$$

where $\mathbf{D}_{(2)}$ is chosen in the same way as $\mathbf{D}_{(1)}$.

Further approximations may be obtained in a similar manner. Defining $\mathbf{L}^*_{(s)}$ in the same way as $\mathbf{L}^*_{(0(1))}$, we have for $s=2, 3, \dots$ sequences of matrices $\mathbf{M}_{(s)}$, $\mathbf{L}^*_{(s)}$, $\mathbf{L}^*_{(0(s))}$, $\mathbf{D}_{(s)}$ satisfying

$$\begin{aligned} \mathbf{M}_{(s)} &= (\mathbf{L}'\mathbf{L})^{-1}(\mathbf{L}'\mathbf{L}^*_{(0(s-1))})\mathbf{D}_{(s)}, \\ \mathbf{L}^*_{(s)} &= \mathbf{LM}_{(s)}. \end{aligned}$$

These sequences converge to matrices which we shall denote by \mathbf{M} , \mathbf{L}^* , \mathbf{L}^*_0 and \mathbf{D} respectively. We have finally, $\mathbf{L}^* = \mathbf{LM}$, where

$$\mathbf{M} = (\mathbf{L}'\mathbf{L})^{-1}(\mathbf{L}'\mathbf{L}^*_{(0)})\mathbf{D}. \quad (1)$$

III. NON-ITERATIVE PROCEDURE FOR CORRELATED FACTORS

The process just described can be simplified so that the same results may be obtained without iteration. We shall show that, for each value of r , the r th column \mathbf{m}_r of \mathbf{M} is the column vector corresponding to the largest latent root of a certain matrix \mathbf{H}_r . Let us denote by \mathbf{L}_r the matrix formed from \mathbf{L} by replacing all the elements in the i th row of \mathbf{L} (or the i th column of \mathbf{L}') by zeros whenever $q_{ir}=0$. For example, if, with $p=7$ and $k=3$, the r th row of \mathbf{Q}' is

$$[1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1],$$

then \mathbf{L}'_r is of the form

$$\begin{matrix} x & 0 & x & 0 & 0 & x & x \\ x & 0 & x & 0 & 0 & x & x \\ x & 0 & x & 0 & 0 & x & x \end{matrix}$$

where x's denote non-zero elements.

Let \mathbf{l}^*_r and \mathbf{l}^*_{r0} be the r th columns of \mathbf{L}^* and \mathbf{L}^*_{r0} respectively.

Then

$$\mathbf{l}^*_r = \mathbf{L}\mathbf{m}_r,$$

$$\mathbf{l}^*_{r0} = \mathbf{L}_r\mathbf{m}_r.$$

From equation (1) it is clear that \mathbf{m}_r is proportional to

$$(\mathbf{L}'\mathbf{L})^{-1} \mathbf{L}' \mathbf{l}^*_{r0} = (\mathbf{L}'\mathbf{L})^{-1} \mathbf{L}' \mathbf{L}_r \mathbf{m}_r = (\mathbf{L}'\mathbf{L})^{-1} (\mathbf{L}'_r \mathbf{L}_r) \mathbf{m}_r. \quad (2)$$

This is equivalent to

$$\mathbf{H}_r \mathbf{m}_r = \lambda_r \mathbf{m}_r,$$

where \mathbf{H}_r is the $k \times k$ matrix given by

$$\mathbf{H}_r = (\mathbf{L}'\mathbf{L})^{-1} (\mathbf{L}'_r \mathbf{L}_r), \quad (3)$$

and where λ_r is a latent root of \mathbf{H}_r . The procedure described in section II is algebraically equivalent to the usual iterative method of finding the largest latent root of a matrix. Hence λ_r is the largest latent root of \mathbf{H}_r , in general distinct, and \mathbf{m}_r is the corresponding column vector. It can be shown that if the r th column of \mathbf{Q} has exactly $(k-1)$ zeros then $\lambda_r=1$. If there are more than $(k-1)$ zeros, then $0 < \lambda_r \leq 1$. In practice λ_r is usually either unity or fairly near unity.

Equation (2) can be reached more directly by starting with a different principle. For each r let us choose \mathbf{m}_r in such a way as to minimize the sum of squares of those elements of $\mathbf{l}^*_r = \mathbf{L}\mathbf{m}_r$ which correspond in position to zero elements of the r th column of \mathbf{Q} , subject to the sum of squares of all elements of \mathbf{l}^*_r remaining constant. That is, we minimize

$$\mathbf{l}^{*'}_r \mathbf{l}^*_r - \mathbf{l}^{*'}_{r0} \mathbf{l}^*_{r0} = \mathbf{m}'_r (\mathbf{L}'\mathbf{L} - \mathbf{L}'_r \mathbf{L}_r) \mathbf{m}_r$$

subject to

$$\mathbf{l}^{*'}_r \mathbf{l}^*_r = \mathbf{m}'_r (\mathbf{L}'\mathbf{L}) \mathbf{m}_r$$

being held constant. This gives the equation

$$(\mathbf{L}'\mathbf{L} - \mathbf{L}'_r \mathbf{L}_r) \mathbf{m}_r - \mu_r (\mathbf{L}'\mathbf{L}) \mathbf{m}_r = 0,$$

where μ_r is a constant, and this is equivalent to

$$(L'_r L_r) m_r = \lambda_r (L' L) m_r,$$

where $\lambda_r = 1 - \mu_r$. Hence, pre-multiplying by $(L' L)^{-1}$, we have

$$H_r m_r = \lambda_r m_r,$$

where H_r is as given above, by (3).

Since rapid and powerful methods exist for finding the latent roots and vectors of *symmetric* matrices, there is an advantage in obtaining the vectors m_r by the following method. Let T be the lower triangular matrix with positive diagonal elements, such that $TT' = L' L$, and let W_r be the matrix given by

$$W_r = T^{-1} (L'_r L_r) T'^{-1}. \quad (4)$$

Then W_r is a symmetric positive definite matrix having the same latent roots as H_r . Let u_r be the vector corresponding to the largest latent root λ_r . Apart from a scaling factor the vector m_r is then given as $T'^{-1} u_r$.

Let M^* denote a matrix in which the columns m_r have entirely arbitrary scales. Let D be the diagonal matrix with positive elements such that D^2 is equal to the diagonal part of $(M^* M^*)^{-1}$. Then the required transformation matrix is given by $M = M^* D$. The matrix $P = (M' M)^{-1}$ has unit diagonal elements and is the correlation matrix for the new factors.

If the original loading matrix L has been found by the maximum likelihood method (see [7], chapter 2), there is a considerable computational advantage in substituting $V^{-1/2} L$ for L throughout, where V is the diagonal matrix whose elements v_i are the residual variances of the x_i , i.e. the components of variance not due to the common factors. With this method the matrix

$$J = L' V^{-1} L$$

is made diagonal, and hence the W_r matrices may easily be calculated as

$$W_r = J^{-1/2} (L'_r V^{-1} L_r) J^{-1/2}. \quad (5)$$

Apart from its scaling factor the vector m_r is given by $J^{-1/2} u_r$.

IV. ROTATION TO NEW ORTHOGONAL FACTORS

If the new factors are to be orthogonal, L must be post-multiplied by an orthogonal matrix. We shall assume that the columns of Q can be ordered in such a way that, for $r = 1, 2, \dots, k-1$, there are at least $(k-r)$ zeros in the r th column of Q . (The last column need have no zeros.)

In this case it seems most convenient to carry out the factor rotation in $(k-1)$ stages. In the first stage we find L_1 , W_1 , u_1 and m_1 as before. The only difference is that we now standardize m_1 so that $m'_1 m_1 = 1$. We next construct an orthogonal matrix M_1 , of order k , with m_1 as its first column (see section V). Then rotate all k factors so that the loading matrix becomes LM_1 . The first column of this, which is Lm_1 , is then the first column of L^* . Now delete the first column of LM_1 and rename the resulting $p \times (k-1)$ matrix L .

With the new matrix \mathbf{L} we calculate \mathbf{L}_2 as before and hence find \mathbf{W}_2 , which is of order $(k-1)$. We then obtain the vectors \mathbf{u}_2 and \mathbf{m}_2 , each containing $(k-1)$ elements. The latter is standardized so that $\mathbf{m}'_2 \mathbf{m}_2 = 1$. Next we construct an orthogonal matrix \mathbf{M}_2 , of order $(k-1)$, whose first column is \mathbf{m}_2 . This enables us to rotate all factors except the first to obtain a new loading matrix \mathbf{LM}_2 . The first column of this, which is \mathbf{Lm}_2 , gives the second column of \mathbf{L}^* . The remaining columns of \mathbf{LM}_2 represent the loading matrix of the last $(k-2)$ factors at this stage. If $k > 3$ the procedure is continued.

At the r th stage of the process \mathbf{L} and \mathbf{L}_r are $p \times (k-r+1)$ matrices and \mathbf{W}_r is a square matrix of order $(k-r+1)$. The vectors \mathbf{u}_r and \mathbf{m}_r each have $(k-r+1)$ elements and \mathbf{m}_r is standardized so that $\mathbf{m}'_r \mathbf{m}_r = 1$. The matrix \mathbf{M}_r is an orthogonal matrix of order $(k-r+1)$, with first column \mathbf{m}_r , which rotates the last $(k-r+1)$ factors. The r th column of \mathbf{L}^* is given as \mathbf{Lm}_r . At the final stage, when $r=k-1$, the last two factors are rotated and \mathbf{LM}_{k-1} provides the last two columns of \mathbf{L}^* .

V. CONSTRUCTING AN ORTHOGONAL MATRIX WHOSE FIRST COLUMN IS GIVEN

In this section we give a method of constructing an orthogonal matrix \mathbf{U} , of order n , whose first column is given. The method is very simple to programme for a computer.

We denote the r th column of \mathbf{U} by \mathbf{u}_r and its m th element by u_{mr} . We define column vectors $\mathbf{z}_2, \dots, \mathbf{z}_n$, each of n elements, by

$$\mathbf{z}_2 = \{0 \ 1 \ 0 \ \dots \ 0 \ 0\},$$

$$\mathbf{z}_3 = \{0 \ 0 \ 1 \ \dots \ 0 \ 0\},$$

•

•

•

$$\mathbf{z}_n = \{0 \ 0 \ 0 \ \dots \ 0 \ 1\}.$$

We may suppose, without loss of generality, that the rows of \mathbf{U} are so ordered that u_{11} , the first element of \mathbf{u}_1 , is not zero or near-zero. This ensures that \mathbf{u}_1 is not linearly dependent upon $\mathbf{z}_2, \dots, \mathbf{z}_n$, or nearly so.

The vectors \mathbf{z}_r enable us to construct without difficulty successive columns of \mathbf{U} . Suppose that, for $m > 1$, the columns $\mathbf{u}_1, \dots, \mathbf{u}_{m-1}$ already exist, and let \mathbf{u}^*_m be the vector defined by

$$\mathbf{u}^*_m = \mathbf{z}_m - \sum_{r=1}^{m-1} u_{mr} \mathbf{u}_r. \quad (6)$$

Then \mathbf{u}^*_m is orthogonal to $\mathbf{u}_1, \dots, \mathbf{u}_{m-1}$ since, for $s < m$,

$$\mathbf{u}'_s \mathbf{u}^*_m = \mathbf{u}'_s \mathbf{z}_m - u_{ms} = 0.$$

We have also

$$\mathbf{u}^*{}' \mathbf{m} \mathbf{u}^* \mathbf{m} = \mathbf{u}^*{}' \mathbf{m} \mathbf{z}_m = u^*_{mm},$$

where u^*_{jm} denotes the j th element of $\mathbf{u}^* \mathbf{m}$. Hence the vector

$$\mathbf{u}_m = (1/\sqrt{u^*_{mm}}) \mathbf{u}^* \mathbf{m} \quad (7)$$

may be chosen as the m th column of \mathbf{U} since it is orthogonal to $\mathbf{u}_1, \dots, \mathbf{u}_{m-1}$ and satisfies $\mathbf{u}'_m \mathbf{u}_m = 1$.

As \mathbf{u}_1 is known initially, relations (6) and (7) enable us to find successively $\mathbf{u}_2, \dots, \mathbf{u}_n$. It may also be noted that, if $n \geq 3$, some of the elements of \mathbf{U} are zero since, for $1 < m < s$,

$$u^*_{ms} = \mathbf{u}^*{}' \mathbf{s} \mathbf{z}_m = \mathbf{u}^*{}' \mathbf{s} (\mathbf{u}^* \mathbf{m} + \sum_{r=1}^{m-1} u_{mr} \mathbf{u}_r) = 0,$$

and hence also $u_{ms} = 0$.

VI. A NUMERICAL EXAMPLE

Transformation to correlated factors

To illustrate the calculations we shall take \mathbf{L}' to be the matrix given in [7], page 20, which was obtained by the maximum likelihood method of estimation. Its elements l_{ir} are:

$$\begin{array}{ccccccccc} 0.664 & 0.688 & 0.492 & 0.837 & 0.705 & 0.820 & 0.661 & 0.457 & 0.765 \\ 0.322 & 0.248 & 0.304 & -0.291 & -0.314 & -0.377 & 0.397 & 0.294 & 0.428 \\ -0.075 & 0.192 & 0.224 & 0.037 & 0.155 & -0.104 & 0.077 & -0.488 & 0.009 \end{array}$$

The elements v_i of the matrix \mathbf{V} , i.e. the residual variances, are given by the formula

$$v_i = 1 - \sum_r l_{ir}^2.$$

They are:

$$0.4498 \quad 0.4283 \quad 0.6153 \quad 0.2134 \quad 0.3804 \quad 0.1747 \quad 0.3995 \quad 0.4666 \quad 0.2315,$$

where, for example, $0.4498 = 1 - 0.644^2 - 0.322^2 - (-0.075)^2$.

Since the loadings are maximum likelihood estimates we use equation (5) and find \mathbf{W}_1 , \mathbf{W}_2 and \mathbf{W}_3 for a pattern matrix, with correlated factors, given by

$$\mathbf{Q}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

The diagonal matrix $\mathbf{J} = \mathbf{L}' \mathbf{V}^{-1} \mathbf{L}$ has elements (14.9863 3.3649 0.8372).

We find that

$$\mathbf{M} = \begin{pmatrix} (\mathbf{m}_1) & (\mathbf{m}_2) & (\mathbf{m}_3) \\ 0.470 & 0.526 & 0.311 \\ 0.773 & -0.977 & 0.148 \\ 0.906 & -0.106 & -1.115 \end{pmatrix}; \quad \mathbf{P} = \begin{pmatrix} 1.000 & 0.431 & 0.507 \\ 0.431 & 1.000 & 0.137 \\ 0.507 & 0.137 & 1.000 \end{pmatrix}.$$

The final loading matrix \mathbf{L}^* is:

$$\begin{array}{ccccccccc} 0.493 & 0.689 & 0.699 & 0.202 & 0.229 & (0.000) & 0.689 & (0.000) & 0.699 \\ (0.043) & (0.099) & (-0.062) & 0.721 & 0.661 & 0.810 & (-0.048) & (0.005) & (-0.017) \\ 0.338 & (0.037) & (-0.052) & 0.176 & (0.000) & 0.315 & 0.178 & 0.730 & 0.291 \end{array}$$

Rotation to new orthogonal factors

In rotating to new orthogonal factors we change \mathbf{Q}' so that what were rows 2 and 3 now become rows 1 and 2 respectively, and row 3 now consists entirely of unities. \mathbf{W}_1 is again found from equation (5). For \mathbf{W}_2 it is necessary to use equation (4) with $\mathbf{L}'_2 \mathbf{V}^{-1} \mathbf{L}_2$ in place of $\mathbf{L}'_2 \mathbf{L}_2$ and with $\mathbf{T}\mathbf{T}' = \mathbf{L}'\mathbf{V}^{-1}\mathbf{L}$, since the latter matrix is, at this stage, no longer diagonal. The rotation matrices are:

$$\mathbf{M}_1 = \begin{pmatrix} (\mathbf{m}_1) \\ 0.471 & 0.860 & 0.196 \\ -0.877 & 0.481 & 0.000 \\ -0.095 & -0.173 & 0.980 \end{pmatrix}; \quad \mathbf{M}_2 = \begin{pmatrix} (\mathbf{m}_2) \\ 0.492 & 0.871 \\ -0.871 & 0.492 \end{pmatrix}.$$

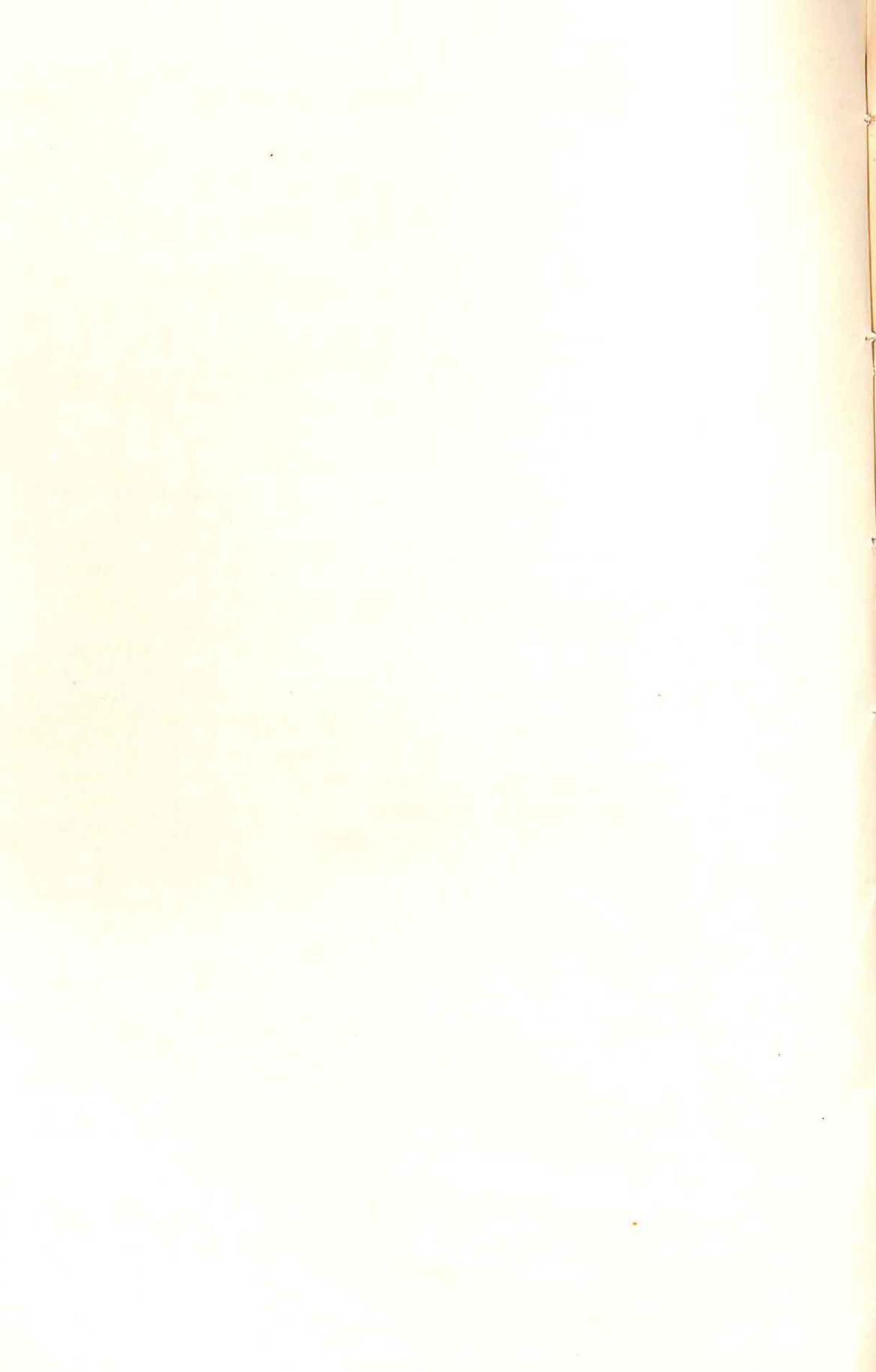
Due to the change in \mathbf{Q}' , the column \mathbf{m}_1 here corresponds, with a different standardization, to \mathbf{m}_2 in the correlated factor case.

The final loading matrix \mathbf{L}^* is:

$$\begin{array}{ccccccccc} (0.037) & (0.088) & (-0.056) & 0.646 & 0.593 & 0.727 & (-0.044) & (0.004) & (-0.016) \\ 0.315 & (0.053) & (-0.014) & 0.108 & (-0.041) & 0.216 & 0.189 & 0.644 & 0.287 \\ 0.671 & 0.749 & 0.618 & 0.597 & 0.515 & 0.501 & 0.750 & 0.347 & 0.828 \end{array}$$

REFERENCES

- [1] AHMAVAARA, Y. (1954). Transformational analysis of factorial data. *Annals of the Finnish Academy of Sciences, Series B*, 88, No. 2, Helsinki, 1-43.
- [2] ANDERSON, T. W. and RUBIN, H. (1956). Statistical inference in factor analysis, *Proc. Third Berkeley Symposium*, 5, 111-150.
- [3] HENDRICKSON, A. E. and WHITE, P. O. (1964). Promax: a quick method for rotation to oblique simple structure, *Brit. J. statist. Psychol.*, 17, 65-70.
- [4] HOWE, W. G. (1955). Some contributions to factor analysis, *USAEC Rep. ORNL-1919*.
- [5] KAISER, H. F. (1958). The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, 23, 187-200.
- [6] LAWLEY, D. N. (1958). Estimation in factor analysis under various initial assumptions, *Brit. J. statist. Psychol.*, 11, 1-12.
- [7] LAWLEY, D. N. and MAXWELL, A. E. (1963). *Factor Analysis as a Statistical Method*, Butterworth & Co., London and Washington.



TESTING DIFFERENCES BETWEEN RELIABILITY COEFFICIENTS¹

By WALTER KRISTOF

University of Giessen, Germany

In the following paper an attempt is made to derive a working formula, based on the method of maximum likelihood, for testing the significance of the difference between two reliability coefficients. It is assumed that both tests have been administered to the same group, and can be split into equivalent halves. A worked example is appended by way of illustration.

I. INTRODUCTION AND PROBLEM

When we want to decide whether two tests are equally reliable an adequate statistical method is needed. Since confidence intervals are available for reliability coefficients (Kristof, 1963), this decision can be made if each test is administered to a different group of subjects. But when the same group of subjects is employed twice, additional information is available and should be used. Accordingly we shall restrict ourselves to the case in which two tests are given to the same subjects. A significance test based on the likelihood ratio will be derived which will provide a method of testing the homogeneity of two reliability coefficients. The sample size N is assumed to be not too small.

This approach treats reliability as an invariant property of the test and thus belongs to classical reliability theory which uses the concept of parallel measures. This concept is equivalent to the true-score-plus-error assumption. Accordingly we define reliability as that portion of the total variation observed which is attributable to the variation of the true measures, and disregard previous extensions of the reliability concept (cf. Burt (1955), Mahmoud (1955), Cronbach, Rajaratnam and Gleser (1963) and other writers). In particular we shall consider only one source of variation, namely, the persons tested.

A further assumption is that each test has been split into two equivalent halves with equal variances. Scores on the halves will be denoted by x_1 and x_2 for the first test, and x_3 and x_4 for the second. Splitting tests into more than two parts will not be considered. We shall also assume that scores on the halves have a joint normal distribution. A test by Walker and Lev (1953, p. 190) may then be used to check the equality of the variances of the corresponding halves. A more adequate method, however, would be Votaw's (1948) test of compound symmetry.

It is convenient to perform a transformation of the variables:

$$\begin{aligned}y_1 &= x_1 + x_2, & y_2 &= x_1 - x_2, \\y_3 &= x_3 + x_4, & y_4 &= x_3 - x_4.\end{aligned}\tag{1}$$

¹ This investigation has been supported by the U.S. Office of Naval Research under Contract Nonr-2752(00).

Let σ_{jk} denote the covariance of y_j and y_k . Then the variance-covariance matrix will be†

$$(\sigma_{jk}) = \begin{bmatrix} \sigma_{11} & 0 & \sigma_{13} & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ \sigma_{31} & 0 & \sigma_{33} & 0 \\ 0 & 0 & 0 & \sigma_{44} \end{bmatrix} \quad (2)$$

with $\sigma_{13} = \sigma_{31}$; and it will have the determinant

$$|\sigma_{jk}| = \sigma_{22}\sigma_{44}(\sigma_{11}\sigma_{33} - \sigma_{13}^2). \quad (3)$$

Let us denote the general element of the matrix which is inverse to (σ_{jk}) by σ^{jk} . It is easily found that

$$\sigma^{11} = \frac{\sigma_{33}}{\sigma_{11}\sigma_{33} - \sigma_{13}^2}, \quad (4a)$$

$$\sigma^{22} = \frac{1}{\sigma_{22}}, \quad (4b)$$

$$\sigma^{33} = \frac{\sigma_{11}}{\sigma_{11}\sigma_{33} - \sigma_{13}^2}, \quad (4c)$$

$$\sigma^{44} = \frac{1}{\sigma_{44}}, \text{ and} \quad (4d)$$

$$\sigma^{13} = \sigma^{31} = -\frac{\sigma_{13}}{\sigma_{11}\sigma_{33} - \sigma_{13}^2}. \quad (4e)$$

These relations will be used later. Completely analogous equations express σ_{jk} in terms of σ^{jk} .

Let us now express the null hypothesis H_0 in mathematical symbols. Since σ_{22} represents the error variance of the first test and σ_{44} of the second, the reliability of the first will be $(\sigma_{11} - \sigma_{22})/\sigma_{11}$ and of the second test $(\sigma_{33} - \sigma_{44})/\sigma_{33}$, reliability being defined as the ratio of true and total variance. Equating both reliabilities, we have

$$H_0: \sigma_{11}\sigma_{44} - \sigma_{22}\sigma_{33} = 0; \quad (5a)$$

and in terms of the σ^{jk} 's we obtain

$$H_0: \sigma^{11}\sigma^{44} - \sigma^{22}\sigma^{33} = 0. \quad (5b)$$

II. MAXIMUM LIKELIHOOD ESTIMATION WITHOUT RESTRICTION

Since H_0 includes second moments only we resort to the Wishart distribution of the unbiased empirical second moments s_{jk} as a basis for the maximum-likelihood ratio test. The density of the Wishart distribution (the likelihood) is

$$L = K |\sigma_{jk}|^{-\frac{N-1}{2}} \exp \left\{ -\frac{N-1}{2} \sum_{j,k=1}^4 \sigma^{jk} s_{jk} \right\} \quad (6)$$

where K is a constant not involving the σ_{jk} 's.

† σ here denotes variance and not standard deviation.

First we derive the maximum-likelihood estimators $\hat{\sigma}_{jk}$ of σ_{jk} when no restrictions are imposed. The quantity to be differentiated may be written

$$q = \ln |\sigma_{jk}| + \sum_{j,k=1}^4 \sigma^{jk} s_{jk}, \quad (7a)$$

or, in terms of the σ^{jk} 's solely,

$$q = -\ln |\sigma^{jk}| + \sum_{j,k=1}^4 \sigma^{jk} s_{jk}; \quad (7b)$$

\ln designates a natural logarithm. It is convenient to differentiate with respect to the σ^{jk} 's rather than with respect to the σ_{jk} 's. Thus we obtain in general

$$\frac{\partial q}{\partial \sigma^{jk}} = (2 - \delta_{jk}) (-\sigma_{jk} + s_{jk}) \quad (8)$$

where δ_{jk} is the Kronecker symbol. Setting these derivatives equal to zero we obtain

$$\hat{\sigma}_{jk} = s_{jk}. \quad (9)$$

For the sake of clarity let us write in detail

$$\hat{\sigma}_{11} = s_{11}, \quad (10a)$$

$$\hat{\sigma}_{22} = s_{22}, \quad (10b)$$

$$\hat{\sigma}_{33} = s_{33}, \quad (10c)$$

$$\hat{\sigma}_{44} = s_{44}, \quad (10d)$$

$$\hat{\sigma}_{13} = s_{13}. \quad (10e)$$

Thus the desired maximum-likelihood estimators have been found.

An essential assumption of these and subsequent derivations is that the tests under consideration do not have very low reliabilities. In that case we need not be concerned about difficulties arising from the additional restrictions $\sigma_{22} \leq \sigma_{11}$ and $\sigma_{44} \leq \sigma_{33}$, which are not implied by the Gramian character of (σ_{jk}) .

III. MAXIMUM LIKELIHOOD ESTIMATION UNDER H_0

As a next step, maximum-likelihood estimators $\tilde{\sigma}_{jk}$ of σ_{jk} under the restrictive hypothesis H_0 are required. Employing a Lagrange multiplier λ , the quantity to be differentiated can be written

$$q = -\ln |\sigma^{jk}| + \sum_{j,k=1}^4 \sigma^{jk} s_{jk} + \lambda (\sigma^{11}\sigma^{44} - \sigma^{22}\sigma^{33}). \quad (11)$$

Setting all partial derivatives equal to zero (note $\sigma^{13} = \sigma^{31}$), we have six equations with six independent unknowns:

$$-\tilde{\sigma}_{11} + s_{11} + \lambda \tilde{\sigma}^{44} = 0, \quad (12a)$$

$$-\tilde{\sigma}_{22} + s_{22} - \lambda \tilde{\sigma}^{33} = 0, \quad (12b)$$

$$-\tilde{\sigma}_{33} + s_{33} - \lambda \tilde{\sigma}^{22} = 0, \quad (12c)$$

$$-\tilde{\sigma}_{44} + s_{44} + \lambda \tilde{\sigma}^{11} = 0, \quad (12d)$$

$$-2\tilde{\sigma}_{13} + 2s_{13} = 0, \quad (12e)$$

$$\tilde{\sigma}_{11}\tilde{\sigma}_{44} - \tilde{\sigma}_{22}\tilde{\sigma}_{33} = \tilde{\sigma}^{11}\tilde{\sigma}^{44} - \tilde{\sigma}^{22}\tilde{\sigma}^{33} = 0. \quad (12f)$$

Since eqn. (12e) does not involve λ , the result

$$\tilde{\sigma}_{13} = \hat{\sigma}_{13} = s_{13} \quad (13)$$

is at once obtained. The remaining five equations determine the entire solution. Somewhat unwieldy cubic equations result for each essential unknown. Thus it seems to be computationally simpler to solve for one unknown, say $\tilde{\sigma}_{11}$, first, and then to find the remaining unknowns from recursive formulae. We will describe one possibility in what follows, other ways being equally feasible.

We find $\tilde{\sigma}_{11}$ from the cubic

$$\begin{aligned} \tilde{\sigma}_{11}^3 - \frac{s_{11}s_{33} + s_{13}^2}{s_{33}} \tilde{\sigma}_{11}^2 + \frac{s_{11}s_{33}(s_{11}s_{44} - s_{22}s_{33}) + 2s_{13}^2(2s_{11}s_{44} + s_{22}s_{33})}{4s_{33}s_{44}} \tilde{\sigma}_{11} \\ - \frac{s_{11}s_{13}^2(s_{22}s_{33} + s_{11}s_{44})}{4s_{33}s_{44}} = 0. \end{aligned} \quad (14)$$

Then $\tilde{\sigma}_{33}$ can be obtained from

$$\tilde{\sigma}_{33} = \frac{s_{33}\tilde{\sigma}_{11}}{2\tilde{\sigma}_{11} - s_{11}}, \quad (15)$$

and subsequently

$$\tilde{\sigma}_{44} = \frac{s_{44}\tilde{\sigma}_{11} + s_{22}\tilde{\sigma}_{33}}{2\tilde{\sigma}_{11}}, \quad (16)$$

and finally

$$\tilde{\sigma}_{22} = \frac{\tilde{\sigma}_{11}\tilde{\sigma}_{44}}{\tilde{\sigma}_{33}}. \quad (17)$$

Thus all $\tilde{\sigma}_{jk}$'s may be calculated from the data. Instead of eqn. (14) we may perhaps, for purposes of computation, use the more convenient form:

$$\begin{aligned} \tilde{\sigma}_{11}^3 - \left(s_{11} + \frac{s_{13}^2}{s_{33}}\right)\tilde{\sigma}_{11}^2 + \frac{1}{4} \left[s_{11}^2 - \frac{s_{11}s_{22}s_{33}}{s_{44}} + 2s_{13}^2 \left(2\frac{s_{11}}{s_{33}} + \frac{s_{22}}{s_{44}}\right)\right] \tilde{\sigma}_{11} \\ - \frac{s_{11}s_{13}^2}{4} \left(\frac{s_{22}}{s_{44}} + \frac{s_{11}}{s_{11}}\right) = 0. \end{aligned} \quad (18)$$

If we introduce the empirical correlation for the total tests, r say, and bear in mind the identity $s_{13}^2 = s_{11}s_{33}r^2$, we may then write

$$\tilde{\sigma}_{11}^3 - s_{11}(1+r^2)\tilde{\sigma}_{11}^2 + \frac{1}{4}s_{11}[s_{11}(1+4r^2) - t(1-2r^2)]\tilde{\sigma}_{11} - \frac{1}{4}s_{11}^2r^2(s_{11}+t) = 0. \quad (19)$$

Here the abbreviation $t = s_{22}s_{33}/s_{44}$ is used.

Some remarks concerning the solution of the cubic will be in order. There is always a positive real root, as is easily inferred from the form of the cubic. This root will be unique (see Kendall and Stuart, 1961, p. 52). To find the root different methods are available. We may proceed graphically, or we may use iterative approximation, gradually approaching the root to any desired degree of accuracy. For this purpose Newton's rule and the "regula falsi" might be helpful. Provided the coefficients in the cubic do not contain too many digits, Cardano's formula (see Burington, 1958, p. 8) can be used to find the root algebraically.

It is advisable to determine as many digits of the root as possible since even small deviations from the true value may seriously affect the whole solution.

IV. THE LIKELIHOOD RATIO TEST

The likelihood ratio l for testing H_0 is

$$l = \frac{L(\tilde{\sigma}_{11}, \dots, \tilde{\sigma}_{44})}{L(\hat{\sigma}_{11}, \dots, \hat{\sigma}_{44})} = \frac{L(\tilde{\sigma}_{11}, \dots, \tilde{\sigma}_{44})}{L(s_{11}, \dots, s_{44})}. \quad (20)$$

By means of eqns. (4) and (10) it is easily verified that

$$\sum_{j,k=1}^4 \hat{\sigma}^{jk} s_{jk} = 4. \quad (21)$$

Analogously we find after some manipulation of eqns. (4) and (12), that

$$\sum_{j,k=1}^4 \tilde{\sigma}^{jk} s_{jk} = 4. \quad (22)$$

Thus the likelihood ratio becomes

$$l = \left(\frac{|s_{jk}|}{|\tilde{\sigma}_{jk}|} \right)^{\frac{N-1}{2}}. \quad (23)$$

In accordance with the general theorem for the likelihood ratio, namely, that in large samples the quantity $-2 \ln l$ is distributed approximately as χ^2 with the number of degrees of freedom equal to the number of restrictions imposed upon the maximization problem under H_0 , we conclude that the quantity

$$\chi^2 = \ln \left(\frac{|\tilde{\sigma}_{jk}|}{|s_{jk}|} \right)^{N-1} \quad (24)$$

is distributed as χ^2 with $df = 1$; or, written out in fuller detail

$$\chi^2 = (N-1) \ln \frac{\tilde{\sigma}_{22}\tilde{\sigma}_{44}(\tilde{\sigma}_{11}\tilde{\sigma}_{33} - s_{13}^2)}{s_{22}s_{44}(s_{11}s_{33} - s_{13}^2)}. \quad (25)$$

Using common logarithms (lg) and resolving the fraction, we may also write

$$\chi^2 = 2.3026(N-1) \cdot [\lg \tilde{\sigma}_{22} + \lg \tilde{\sigma}_{44} + \lg (\tilde{\sigma}_{11}\tilde{\sigma}_{33} - s_{13}^2) - \lg s_{22} - \lg s_{44} - \lg (s_{11}s_{33} - s_{13}^2)]. \quad (26)$$

This is perhaps the simplest form for purposes of computation. Thus the desired maximum-likelihood ratio test of homogeneity of two empirical reliability coefficients has been obtained.

The foregoing derivation assumes that for both tests the total scores and the difference scores on the halves are available. Sometimes, however, it may be more convenient to use the empirical variances of the totals and the halves only. In this case we may compute the required quantities s_{jk} with the help of the following equations

$$s_{11} = s_{x_1+x_2}^2; \quad (27a)$$

$$s_{22} = 2(s_{x_1}^2 + s_{x_2}^2) - s_{x_1+x_2}^2, \quad (27b)$$

$$s_{33} = s_{x_3+x_4}^2, \quad (27c)$$

$$s_{44} = 2(s_{x_3}^2 + s_{x_4}^2) - s_{x_3+x_4}^2, \quad (27d)$$

$$s_{13} = s_{x_1+x_2, x_3+x_4}; \quad (27e)$$

empirical unbiased variances and covariances being written in the usual fashion.

V. A NUMERICAL EXAMPLE

By way of illustration the method developed in this paper will be applied to numerical data taken from an earlier work by Lord (1957)†.

Taking the test scores of $N=649$ examinees the following covariance matrix was obtained:

$$\begin{bmatrix} 86.3979 & 57.7751 & 56.8651 & 58.8986 \\ 57.7751 & 86.2632 & 59.3177 & 59.6683 \\ 56.8651 & 59.3177 & 97.2850 & 73.8201 \\ 58.8986 & 59.6683 & 73.8201 & 97.8192 \end{bmatrix}$$

In the matrix the first two variables are parallel halves of a vocabulary test administered under very liberal time limits. The last two variables are parallel halves of a vocabulary test constructed so as to be as nearly equivalent as possible to the halves of the first test except that the time of administration was so short that only two per cent of the examinees completed each half.

From the above data the reliability coefficients of both tests are easily found. They are $r_{tt}^{(1)}=0.80$, $r_{tt}^{(2)}=0.86$, the superscripts indicating the tests. We want to decide the question whether $r_{tt}^{(1)}$ and $r_{tt}^{(2)}$ are significantly different, i.e., whether the time conditions have influenced the reliabilities.

With the aid of eqns. (27), the matrix (s_{jk}) is determined:

$$(s_{jk}) = \begin{bmatrix} 288.2113 & 0 & 234.7497 & 0 \\ 0 & 57.1109 & 0 & 0 \\ 234.7497 & 0 & 342.7444 & 0 \\ 0 & 0 & 0 & 47.4640 \end{bmatrix}$$

In this matrix all the entries corresponding to a $\tilde{\sigma}_{jk}=0$ are set equal to zero. Next we find

$$r^2 = \frac{s_{13}^2}{s_{11}s_{33}} = 0.5578645$$

and

$$t = \frac{s_{22}s_{33}}{s_{44}} = 412.4061.$$

Now eqn. (19) becomes

$$\tilde{\sigma}_{11}^3 - 448.9941\tilde{\sigma}_{11}^2 + 70544.76\tilde{\sigma}_{11} - 8116553 = 0.$$

The root of this cubic was found to be

$$\tilde{\sigma}_{11} = 304.940.$$

Then using eqns. (15), (16), and (17) the remaining $\tilde{\sigma}_{jk}$'s were determined. Their matrix is

$$(\tilde{\sigma}_{jk}) = \begin{bmatrix} 304.9400 & 0 & 234.7500 & 0 \\ 0 & 50.8282 & 0 & 0 \\ 234.7497 & 0 & 324.9193 & 0 \\ 0 & 0 & 0 & 54.1584 \end{bmatrix}$$

The entries corresponding to a $\sigma_{jk}=0$ are placed equal to zero.

† I am indebted to Dr. Lord for permission to use his data and for valuable advice.

Now we find

$$\tilde{\sigma}_{11}\tilde{\sigma}_{33} - s_{13}^2 = 43973.47$$

and

$$s_{11}s_{33} - s_{13}^2 = 43675.39.$$

Taking logarithms and using eqn. (26) we end with

$$\chi^2 = 14.39.$$

With one degree of freedom this value is significant at the $P < 0.02$ per cent level. Thus it has been proved that the different time conditions under which the tests were given do influence the test reliabilities, since the reliability of the second test is significantly higher than that of the first.

REFERENCES

- BURINGTON, R. S. (1958). *Handbook of mathematical tables and formulas*. Sandusky, Ohio: Handbook Publishers.
- BURT, C. (1955). Test reliability estimated by analysis of variance. *Brit. J. statist. Psychol.*, 8, 103-118.
- CRONBACH, L. J., RAJARATNAM, N. and GLESER, GOLDINE C. (1963). Theory of generalizability: a liberation of reliability theory. *Brit. J. statist. Psychol.*, 16, 137-163.
- KENDALL, M. G. and STUART, A. (1961). *The advanced theory of statistics*. Vol. 2. New York: Hafner.
- KRISTOF, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, 28, 221-238.
- LORD, F. M. (1957). A significance test for the hypothesis that two variables measure the same trait except for errors of measurement. *Psychometrika*, 22, 207-220.
- MAHMOUD, A. F. (1955). Test reliability in terms of factor theory. *Brit. J. statist. Psychol.*, 8, 119-135.
- VOTAW, D. F., Jr. (1948). Testing compound symmetry in a normal multi-variate distribution. *Ann. math. Statist.*, 19, 447-473.
- WALKER, HELEN M. and LEV, J. (1953). *Statistical inference*. New York: Holt.

THE DETECTION OF CHANGE AND THE PERCEPTUAL MOMENT HYPOTHESIS¹

By T. SHALLICE

University of Manchester

This paper examines Stroud's theory that the perceptual input is quantized in time (the perceptual moment hypothesis). A model of the detection of change based on the theory is compared with other models derived from statistical quality control methods; in particular those using simple and geometric moving averages. They are compared theoretically to see which entails the least computational load for the brain. Empirical comparisons are made with data from loudness thresholds, brightness thresholds, and the perception of causality: all cases in which there is little change in stimulus intensity. The perceptual moment hypothesis is found to be the most satisfactory. Finally, some of the assumptions of the perceptual moment model are examined in more detail and the model is extended to deal with situations in which the intensity of stimulation alters greatly.

I. INTRODUCTION

Stroud (1949, 1955) put forward the theory that "with respect to input-output relations (of man) critically dependent on man's awareness of what is going on, . . . the most efficient description of the handling of these input-output relations is in terms of the psychological time T , which represents t (physical time) in these operations . . . T is not a continuous variable." To put it another way, the brain processes information discontinuously in time, each operation lasting a 'moment'.

Although a considerable amount of evidence has been produced in support of this theory—see, e.g., Stroud (1955), White (1963)—two questions, which would appear to be very important when appraising it, do not seem to have been answered: (i) Could alternative theories based on the assumption that the brain operates in a continuous manner also provide a satisfactory account of the evidence? (ii) Why should the brain operate in this discontinuous fashion? The second question may appear incongruous, but it seems implausible that there should be no evolutionary or other biological reason why the brain should operate discontinuously rather than continuously. In other words, there must be functional arguments which bear on this problem.

Since most of the evidence concerns perceptual aspects of the psychological process we will consider only the 'perceptual moment' hypothesis, i.e., that

¹ I should like to thank R. J. Audley for much advice and encouragement, and in particular for suggesting the idea which became the framework of the paper. I am also grateful to the Medical Research Council for a scholarship which supported part of the work

sensory input is summated over discrete moments of time. This is a weaker and more tractable form of the hypothesis of a psychological moment.

In order to provide a framework within which we can try to answer the two questions we will adopt a functional and cybernetic approach—considering theoretically how an organism can perform a simple but very necessary task. The task to be considered is that of detecting a change in the environment, an analogous task to that of a quality control statistician trying to detect a change in the quality of the output of a manufacturing process. This conceptual approach bears certain similarities to the comparison of the human observer with an ideal observer which has received prominence in Signal Detection Theory and also to the use by Stone (1960) and Laming (1962) of an optimal statistical decision procedure as a model for human performance in choice reaction time experiments. The similarity lies in demonstrating the economical and optimal properties of any means of carrying out a given function. Here, several alternative hypotheses will be compared from this point of view.

The analogy with statistical quality control mentioned above was suggested by Audley (1963). He was interested in how quickly an organism can detect a change in the intensity of a stimulus as a particular case of the reaction time situation. His model comprised a simplified artificial sense organ connected to a detector mechanism by, in the simplest case, one neural channel. If $x_1, x_2 \dots x_m$ is the sequence of inputs which arrive at the detector in successive instants of time up to the m th, the detector has the function of deciding whether they should all be considered as samples from a common distribution representing the background illumination, or whether from some instant onwards they should be regarded as coming from a new and different distribution due to the presence of a signal to which the subject must respond. The detector is in a cybernetically similar situation to that of the statistician engaged in quality control, who has to discover when faults occur in a manufacturing process by observing its output. In the manufacturing analogy, the quicker the change in output is detected, the sooner the process can be repaired. On the other hand a false alarm can be expensive. For the organism, the quicker a change in the environment is detected, the sooner the appropriate action can be taken, but again a false alarm wastes time and attention. We shall make the assumption that the schemes that statisticians have derived for quality control can provide us with a set of models of how the brain detects change which will have greater *a priori* likelihood of being correct than other models, since they have proved their value in a cybernetically similar situation. Some of the models entail a form of the perceptual moment hypothesis, others entail a continuous processing of information.

II. THE QUALITY CONTROL ANALOGY

In a manufacturing process it is normally possible at times t_i ($i = 1, 2, \dots, n$) to calculate the value of a variable y_i , which is related to the quality of the output, x_i .

If the distribution from which y_i is drawn changes at any time, this indicates that there has been some, usually adverse, change in the manufacturing process. If we know Y , the original distribution of the y_i , it is possible to construct what is known as a control chart (see Figure 1), from which it is easy to see whether the chance that a particular y_m came from Y is less than a fixed small amount. If it is, then the manufacturing process is examined since it is very likely that a fault in the process has occurred. The variable y_i can be calculated in many different ways from the actual observations x_i . In practice, only a few ways of calculating y_i from x_i are used, each having virtues important in particular situations. For more details of quality control methods see Grant (1952) and Page (1954).

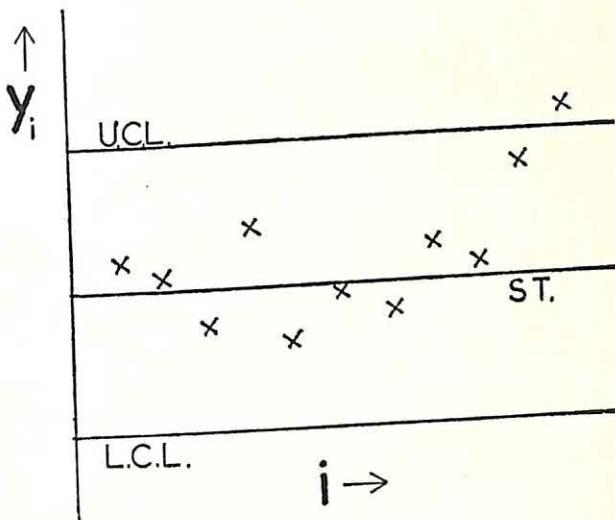


FIGURE 1. A typical control chart. ST represents the average acceptable quality of the output. U.C.L. and L.C.L. are the upper and lower control limits. Remedial action is taken if an observation falls outside these limits, e.g., the last point shown.

These few methods, favoured by statisticians, will provide the bases of the models to be considered, but before going into the details of them there are certain general points which are relevant to the analogy:

(i) What corresponds in the brain to the statisticians' observation? For the organism the input is arriving continuously in time, whilst in quality control the observations are almost always sampled at discrete intervals, e.g., in the manufacture of screws, batches of screws will be selected at particular places in the total output, and their diameters measured. Some quality control methods can be simply adapted to deal with continuously sampled material, others cannot. If a method can be so adapted we will do so for three reasons: to use a discontinuous sampling method, when a continuous method can be applied, loses information; the brain seems, at least superficially, to operate continuously;

and finally it is better to make assumptions against the hypothesis being examined, rather than in favour of it.

There seem to be two possible analogies to the statisticians' observation. These can be contrasted in terms of what happens to impulses in a single nerve fibre. According to one notion there is a summation of the input in an accumulator with a fixed time-span or rate of decay. The second, inherently discontinuous, uses as the variable the length of the interval between two impulses—the principle of Pulse Interval Modulation discussed by MacKay and McCulloch (1952). We will discard the second possibility for reasons given in the Discussion, but basically because it would impose a much greater computational load on the brain. Proceeding with an accumulator model, we shall make a further assumption that in order to deal with the perception of movement a spatial averaging process takes place in the accumulator. Again a fuller discussion of this point will be given later, but basically the reason is to lessen the brain's computational load. By a 'spatial averaging process' is meant one in which the output of the accumulator contains only the average position of an object, and no further details of the transient positions it was in during the time-span of the accumulator. (Throughout the paper the length of time that information takes to reach the accumulator from an object will be ignored—in the cases to be considered it is constant.)

(ii) There seem to be two perceptual situations for which the analogy is suitable—the judgement of change and the perception of change. That the two are different is illustrated by the experiments of Grindley (1936) and Drew (1937). They gradually increased or decreased the intensity of a stimulus. When the subjects reported a change they made one of two types of response. Sometimes they said they perceived the change itself, occasionally not even knowing which way it had changed. At other times they said that they knew the stimulus had changed, but had never perceived it changing. An everyday example of this same dichotomy is the difference between the detection of the movement of the second hand of a clock and that of the hour hand.

One important difference between the two situations is that the judgement of change requires that the subject be concentrating on the stimulus, while the perception of change usually does not. In fact it seems likely that the perception of change comprises an important factor in shifts of attention: if that part of the environment which is not being attended to does not change, it is not normally necessary for the organism to shift its attention to it.

The functions of the two types of detection are different. In the perception of change, change may take place along many different dimensions, and it is especially important to detect large changes quickly. For the judgement of change only a few dimensions are relevant and accuracy is as important as speed. These different requirements will probably lead to different types of mechanisms. This paper will be primarily concerned with models for the perception of change which will be treated as a process which detects change for the perceptual mechanisms mediating recognition and shifts of attention.

(iii) Each quality control system can provide the basis of a model for the way the brain perceives change. One of the models which has the most economical requirements for storage and computing space is presumably more likely to be the correct one, since it would be likely to have evolved or be within the learning capacity of an organism. For instance, some models require that all the incoming data be analysed along separate dimensions before they can be investigated. Others treat them in an unanalysed form, and these seem much more likely to be found since they require so much less computing space.

(iv) In quality control problems the statistician has a fixed standard against which he tests each observation. The functionally corresponding neural process, which we shall call the Change Detector, must deal with a much more complex position, since it must always be creating a new standard, at least on every occasion when a change is detected. Also it must allow for the changes in input due to movements of the organism itself, which are of no use for detecting changes in the environment. Hence new standards must be created frequently. How the new standards are created is not a vital issue when comparing the different models, so this question will be taken up more fully in the Discussion, together with its relation to the classic problem of the Stability of the Visual World. However, one point will be mentioned which is more immediately relevant; the standard must be based on the output of the accumulator at least a certain time before the output with which it is being compared. This is necessary if change is to be detected at all.

(v) It will be assumed, until disproved, that one model is applicable for all modalities, except those where there is any autocorrelation in the input, since then the analogy with Quality Control breaks down. This, in particular, rules out the consideration of the perception of changes of sound quality for notes of high or medium pitch.

III. THE MODELS

(i) *Perceptual Moment Model* In the simple control chart method one puts $y_i = x_i$, i.e., each observation forms the variable which is used as the basis of a decision. The obvious way of creating a model is to choose as the variable x_i the output of the accumulator at the end of each perceptual moment, the accumulator being cleared after each output. (x_i might then be taken as corresponding to the number of impulses that have arrived from a small number of receptor elements.) The standard will presumably be created from the output of the accumulator (or perhaps the Change Detector) of the previous moment. In this model the entire input need not be analysed into dimensions before it reaches the Detector, and for this reason it would not seem to require any complex neural system.

(ii) *Moving Average Model* Whilst the simple control chart is good for detecting those faults in a manufacturing process which result in a large change

in the average output, \bar{x} , it is not so good at detecting a small change in \bar{x} or a slow trend away from the standard. A more sensitive test is obtained by putting

$$y_i = \sum_{r=n(i-1)+1}^{ni} x_r,$$

but this does not lead to a new model, since it is equivalent to making the discrete periods of accumulation larger. A slightly more complex test considers runs of trials in which the variable lies on one side of the standard. This is inherently discontinuous, and difficult to adapt to a brain model. It has certain similarities, but no advantages over the next type to be considered—the moving average model.

In the moving average method we make

$$y_i = \sum_{r=0}^n x_{i-r}.$$

If this method is not to depend on the integration of input over perceptual moments it needs to be transformed into a continuous method in the form:

$$y_T = \int_{T-P}^T x_t dt,$$

where $x_t = 1$ if a pulse arrives at time t

$= 0$ otherwise, and P is the period of the moving average.

The difficulty with the Moving Average Model is that the actual instant when each pulse arrives has to be remembered for P units of time, so that the value of each pulse can be subtracted from the accumulator after it has been stored for that period. One possible way of avoiding this difficulty is to use an 'over-writing' method in which the input is put on to the neural analogue of a rotating store, which is rotating with period P . The contents of the rotating store would be continuously summed. This would probably require that the input from a particular nerve fibre would go to a different place depending upon the phase of the moving average cycle, and an 'eraser' should exist which would clear that part of the store just about to be used. This possibility will not be more fully explored here, since the important point is that a moving average model would have to be much more neurally complex than the perceptual moment model, and so, other things being equal, would be much less likely to be found in the brain.

(iii) *Geometric Moving Average Model (G.M.A.)* Another method, suggested by Roberts (1959), uses a geometric moving average so that the variable for detecting change is

$$y_i = \sum_{r=0}^{\infty} \lambda^r x_{i-r}, (\lambda < 1).$$

In the present context it is much more natural to make the model continuous and consider

$$y_T = \int_0^T \lambda^{T-t} x_t dt.$$

An important feature of this model is that:

$$y_{T+DT} = \lambda^{DT} y_T + \int_T^{T+DT} x_t dt - \epsilon,$$

where $0 < \epsilon < (1 - \lambda^{DT})$ when an impulse arrives in the interval from T to $T + DT$, and otherwise $\epsilon = 0$. As $DT \rightarrow 0$ the maximum error $\rightarrow 0$ and the number of times the error occurs remains constant in any given finite time. So the effect of the errors, ϵ , $\rightarrow 0$ as $DT \rightarrow 0$.

This model corresponds to one with an exponentially decaying accumulator fed by a discontinuous input. It requires as little computing space as the perceptual moment model and seems very plausible neurally, being similar to many psychological theories of brain function.

(iv) *Sequential Sampling Model* Page (1954) discusses the quality control methods considered so far, with the exception of the G.M.A., and explains that as they are based on only a limited number of the more recent observations they do not make use of all the information obtained. He suggests that it would be better to use a scheme based on sequential sampling.

He defines the statistic

$$S_n = \sum_{i=1}^n (x_i - E),$$

where E is the standard, and quality control action is taken if $S_n - \min_{0 \leq i \leq n} S_i \geq h$ (a constant). This will only detect an increase in x_i , and a similar test is needed to detect a decrease. The other tests so far described do these tasks together. If we were to base a model on this test it would be necessary for the stimulus input to be analysed into dimensions before the detection of change along each dimension could take place, in order that the value of the minimum could be ascertained. This makes it totally inferior as a model for the perception of change when compared with models where the raw input can be tested. It also seems to be inherently discontinuous, so to neglect it will if anything count against the perceptual moment hypothesis. While the above objections will allow us reasonably to ignore it as a model for the perception of change, they are much less cogent if it is advanced as a model for the judgement of change. In fact it then gives rise to a model very similar to that proposed by Stone (1960) and elaborated by Laming (1962) for the description of choice reaction time phenomena.

Theoretical Assessment The aim of this section is to provide an assessment of the *a priori* odds in favour of the rival models for the perception of change. The sequential sampling model will be ignored from here on, since its prior chances seem very low. Of the other three, the moving average model is considerably more complex than the other two, which must count against it. The odds on the other two are roughly equal although the perceptual moment model has the advantage that, unlike the others, it provides information about the magnitude of the change as well as detecting change.

Comparisons of the speed and accuracy of the three tests depend upon the value of certain parameters; length of the moment or moving average, and rate of decay of the G.M.A. It seems that for comparable accuracies and for the detection of large changes the moving average is the fastest and the G.M.A. the slowest. However, since the speed of all systems can be increased by changing the parameters it seems unlikely that the small advantage of the Moving Average is enough to compensate for its much greater complexity. For medium change the perceptual moment is probably fastest and for small changes the G.M.A.

These *a priori* considerations seem to make the perceptual moment slightly more likely than the G.M.A., and both considerably more likely than the moving average.

IV. EXPERIMENTAL EVIDENCE

Before discussing the experimental evidence there is one major point that needs to be considered—the values of the parameters of the models. It will be considered in the context of the length of the perceptual moment (D), but these arguments presumably also apply to the other models.

In the literature there is no agreement amongst the different theorists about the length of the moment. The theorists who connect it with the alpha rhythm (Walter, 1950; Wiener, 1948; Murphree, 1954) consider it to be about 100 msec. White (1963), on the basis of subjects' estimates of the number of stimuli in a rapid sequence, deduces that it is about 80 msec. Ansbacher (1944), using the Brown Circle Illusion, deduces that it is about 55 msec, a value quite close to the one we will shortly obtain using Michotte's (1963) data on the perception of causality. If these experiments are comparable, then it seems that the moment length must be variable. Stroud (1955) agrees with such a notion, giving among his postulates the following:

“(IV) $0.05 < D < 0.20$ seconds approximately. Generalization (iv) is based upon experience . . .

(V) t_0, t_1, \dots, t_n (the vector of moment boundaries) is frequently determined by singularities in $f(t)$. I have often found cases where assuming that $D = \text{constant}$ didn't make very good sense—while simple assumptions of synchrony between special point values of the input function and the dates t_0, t_1, \dots , etc., did produce good explanations of the sort we are now considering.”

If the moment length does change, then an obvious candidate for a factor influencing the length of the moment is the intensity of stimulation. If the intensity of stimulation is increased, then the rate at which impulses reach the accumulator will also increase. Hence the number of impulses which will provide an adequate sample for the detection of a change with the same false alarm rate will be reached in a shorter time. So the length of the moment can be decreased. Empirical support for such a decrease in the length of the moment

is that C.F.F., optimal apparent movement, two-flash threshold and the critical period of the Bunsen-Roscoe Law are all greatly affected by a change in the intensity of the stimulus. Since all these phenomena are connected with 'change' it should be expected that they are explicable by the present theory of the perceptual moment.

Assuming that moment length is some function of the intensity of stimulation, then when the level of stimulus intensity changes rapidly the moment length will have to adapt itself to the new level of stimulation. How this might be done is a complex problem, and it will definitely affect the predictions that the theory of moments makes in any situation where the intensity changes rapidly, such as White's (1963) countability experiments. Since the parameters of the other two models are presumably also affected by intensity, the simplest way to discriminate between the models is by using experimental situations in which no great change of intensity occurs. Three such situations will be examined below. The only other known to me—the Brown Circle Illusion examined by Ansbacher (1944)—does not seem to favour any one of the models.

(i) Loudness Thresholds for Low Tones

If there is autocorrelation in the input, then the analogy we have considered would seem to be inadequate since there is an extra source of information in the input which is not being used. This would mean that more efficient methods of detecting change should be available. On the other hand, a device searching for possible periodicities in the input could detect only a limited range of frequencies. The ear is a device of this sort with the basilar membrane acting as a periodicity detector. Yet it seems likely that at low frequencies (less than 300 c.p.s. according to Licklider, 1951) pitch is identified principally by differences in the frequency of neural volleys to the brain and the effect of any autocorrelation detector becomes unimportant. Hence our models are relevant to perception of changes in these low tones.

It might also be wondered whether absolute thresholds should be accounted for by a model of the perception of change. However, Bekesy, whose evidence we will use, says: "In successive trials the sound pressure from the thermophone was reduced by steps of about 10 per cent, and the frequency was slowly raised automatically, starting from 2 c.p.s., until an auditory sensation was aroused." What Bekesy has called the arousal of an auditory sensation must necessitate the perception of a change.

Stroud has already used the nature of the threshold/intensity curve as evidence for the theory of moments, using Bekesy's (1936) discovery that the threshold intensity for tones of low frequency, when plotted against frequency, gives a step function (see Figure 2). Stroud states that "the moment train should 'beat' slowly against the spontaneous frequency of the D's, thus maximizing the likelihood of its being detected at the following frequencies:

$\frac{1}{2D}$, $\frac{1}{D}$, $\frac{2}{D}$, $\frac{3}{D}$, etc." He fits Bekesy's results with $D = 106$ msec and predicts

that the steps should occur at the following frequencies 4.7, 9.4, 18.8, 28.2 and 37.2 c.p.s., which agree very well with the observed frequencies of 4.47, 9.18, 18.3, 28.2 and 38.2 c.p.s.

To relate the data to the models, we will recast Stroud's explanation in physiological terms. It is known that a pure tone of less than 2,000 c.p.s. produces a volley of impulses in the auditory nerve at about the same phase in each cycle, varying only over a range of about 0.13 msec (Davis, 1951). Now the number of impulses per volley, V , will be related to the sound pressure of the tone, P , by some expression like $V = V(P, F)$, which will not depend critically on F , the frequency of the tone.

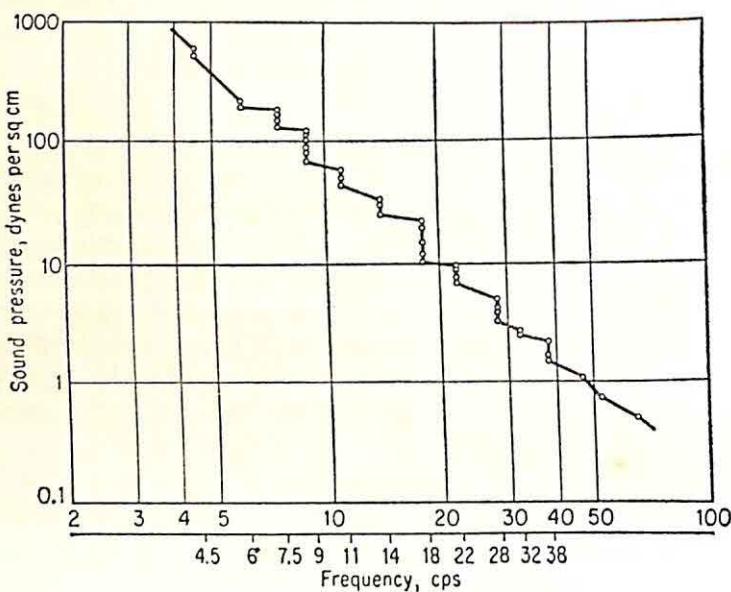


FIGURE 2. Loudness threshold curve at low frequencies (from Bekesy, 1936, *Ann. Physik*, 26, 554–566).

Consider the performance of the perceptual moment model as the frequency of the tone is gradually increased. When the period of the tone is greater than D msec the number of *volleys* arriving in a moment can never be greater than unity, so that the maximum possible number of *impulses* per moment will be $V(P, F)$. When the period is just less than D msec, occasionally there will be two volleys arriving in one moment, so that the maximum number of impulses per moment will now be $2V(P, F)$. For detection to occur the number of impulses per moment should exceed some constant, so that as the period decreases through D msec the threshold sound pressure will have a step. There will be similar steps at $2D$, $3D$, $4D$. These will be decreasing in size because the effect of an additional volley becomes proportionately smaller. The Moving Average

Model will predict similar steps by an almost identical argument. The G.M.A. model, being inherently continuous, seems incapable of explaining the steps. The argument could have been made slightly more rigorous by considering Q , the frequency of hearing, but the extra complexity of the argument brings no extra reward. The conclusion is the same.

Stroud's 'beating' theory did not explain the steps Bekesy found at approximately 6, 7.5, 11, 14, 22 and 32.5 c.p.s. (Figure 2) which are smaller than the main steps. In point of fact these corroborate the moment theory, as it has been shown by Wever, Bray and Willey (1937) that at low frequencies there are normally 2 or 3 volleys per cycle, the smaller ones being seemingly triggered off by the higher harmonics produced by the ear. These would produce small steps in threshold between the larger ones by an extension of the same argument.

It seems that a model which produces some type of discontinuity in the sampling of the sensory input is very strongly supported by Bekesy's results. Hence the G.M.A. model becomes unlikely.

(ii) Brightness Thresholds

The usual explanation of the breakdown of the Bunsen-Roscoe Law for brightness thresholds ($IT = k$ for $T < T_c$) was that temporal quantization occurred at the receptor level either photochemically or due to a fixed neural integration period. This period is triggered by the first quantum absorbed and persists for some critical duration, during which the number of quanta absorbed are counted, and a response is subsequently produced which is in proportion to this number. For various reasons this explanation seems to have gone out of favour and a more continuous view of the activity of the receptor is now envisaged. Mueller (1954), Johnson (1958), Boynton and Siegfried (1962) and Blackwell (1963) have put forward rather similar theories from different experimental approaches. For example, Boynton and Siegfried support the theory that "Every element of stimulus elicits after a latent period, a tiny element of response; the gross response to a stimulus is the sum of the elemental responses to each elemental component of the stimulus."

If this theory is correct, then as Boynton and Siegfried point out, the temporal quantization implied by the Bunsen-Roscoe Law must come at a higher level of the nervous system than the receptor, and one of the detection models we are considering would be an obvious candidate. The discussion given below of how well the three models fit the data will not be rigorous because of various difficulties encountered in this field of study. One of these arises because at a given background and area of flash the chance of the flash being seen, Q , is affected by two stimulus variables—the length of the flash, T , and its intensity, I , and also by a host of subject variables of which the best known is the non-random pattern of subjects' 'yes' and 'no' responses (for references see Bouman, 1961). Another reason is that the three models would require computer simulations before their behaviour could be accurately assessed. However, two approximate predictions can be derived from the models.

The first of these concerns the relation holding between T and I for a fixed value of Q . The G.M.A. model, in which the input of each impulse into the accumulator is treated as a step increase, is extremely difficult to handle mathematically. However, by treating the input as continuous, a fair approximation to its performance can be obtained. Let the momentary level of the accumulator be $g(t)$, the rate of input into it be a (this will be proportional to the intensity of stimulation), and the rate of decay be proportional to $g(t)$, the constant of proportionality being b .

$$\text{Then } dg/dt = a - bg. \\ \therefore bg = a - K e^{-bt},$$

where K is a constant depending upon initial conditions.

In the steady state $g = a/b$.

Now suppose the rate of input changes stepwise at $t=0$ from a to A . Then for $t > 0$, $bg = A - (A-a)e^{-bt}$, since when $t=0$, $g=a/b$. Let the input be N additional photons arriving randomly over T seconds ($\therefore N=IT$), then $T \cdot (A-a) = N$.

$$\therefore b \cdot [g(t)]_{\max} = b \cdot g(T) = a + \frac{N}{T} (1 - e^{-bT}).$$

This continuous model is more appropriate for classical psychophysical methods antedating signal detection theory when thresholds were thresholds. However, we will assume that when $g(t) - g(0) > J$ detection of the signal occurs. (If we wish to consider frequency of seeing curves the obvious approach is to take a given value of N to correspond monotonically to a given value of Q .) Therefore, for a constant chance of seeing we should have

$$N = TJ/(1 - e^{-bT}).$$

This is, however, a very poor approximation to the actual curves of (N, T) . (This is demonstrated in an appendix.) Predictions from the other two models are difficult to obtain, but it is apparent that they, unlike the G.M.A. model, can explain the sharp increase in dN/dT just above $T = 100$ msec.

A second relation that the models could theoretically predict is the relation between N and Q for fixed T . This is in practice very difficult to obtain; but for the perceptual moment model, if the relation between T and N for fixed Q is known, it is reasonably easy to deduce the relation between Q and N for fixed T . Brindley (1960 b) states a theorem, proved in Brindley (1960 a), that "If the flash is long compared with D , a relation of threshold to duration of the form $TI^n = C$ implies a frequency of seeing curve of the form $Q = 1 - e^{-kI^n}$ where k is the positive constant $-[T \ln (1 - Q_0)]/C$."¹ He continued: "The deductions are directly applicable to the data of Bouman and van der Velden (1947), who measured the relation of absolute threshold to area and duration at probability-criterion 0.6, and the frequency of seeing curves at all combinations of four durations and four areas. Bouman and van der Velden found that for small areas the relation of threshold to duration was of the form $TI^2 = c$ in the range

¹ Symbols changed to conform to those used earlier in this paper.

100 msec to 5 sec. The corresponding frequency-of-seeing curves fit the theoretical relation $Q = 1 - e^{-kT^2}$ very well" (p. 192). Hence the perceptual moment model receives some support.

The brightness threshold results are not very conclusive, but taken in conjunction with those emerging from the study of loudness thresholds previously discussed they make the G.M.A. Model implausible. However, they do not greatly help in deciding between the other two models.

(iii) *Perception of Causality*

In Michotte's (1963) book on the Perception of Causality, he describes a large number of experiments in which the subject sees a slit containing a red square, known as 'object A' and some distance away a black square known as 'object B'. Object A moves along the slit at a constant speed and stops when touching object B, after which object B moves off at a possibly different constant speed, and then stops. Various factors such as the speeds and sizes of A and B are varied, and Michotte was interested in whether an impression of causality is reported by the subject.

One of the factors in which Michotte was interested was what happened when an interval was introduced between the end of A's movement and the beginning of B's, the two objects being in contact during the interval. Michotte describes three phenomenological stages in perception as the length of the interval is increased. They are the impressions of 'direct launching', 'delayed launching' and of 'two independent movements'. Launching is the name for the impression of one object 'bumping into' another and setting it in motion. Michotte (p. 92) describes the three stages as: "In the first stage we have the familiar Launching Effect. At the second stage when the interval is greater, the causal impression is still unmistakable, but the Launching Effect involves some 'time-lag'. Object B 'sticks' to Object A, its departure takes place only after some delay. At the third stage, however, the causal impression disappears and is replaced by an impression of successive movements of two objects which form a 'whole'—one which is loosely integrated, there being no internal links".

In order to explain this phenomenon in detail, we will use Michotte's theory of the Perception of Causality as a starting point. Although Michotte's theory is presented in phenomenological terms, there is no great difficulty in translating the arguments into terms congruent with the approach adopted here. The theory of the origin of the Launching Effect is based on many ingenious experiments, and is that: "At the moment of impact, the movement of the motor object (Object A) appears to extend to the projectile (Object B), bringing about its displacement in space. This amounts to saying that from that point on the movement presents a double aspect; we see it both as a continuation of the movement of the motor object before the impact, and as something which brings about a change in the position of the projectile" (p. 140). This 'extension' of the movement of the motor object on to the projectile is called

by Michotte the "ampliation" of the movement. One of the necessary conditions for the impression of causality to occur that Michotte deduces from his theory is that there must be continuity between the two movements. For our present purpose this means that information that movement is occurring must continually be signalled by the Change Detector for launching to be perceived.

In terms of the perceptual moment model this implies that at each moment during the trial, movement, either of *A* or of *B*, must be signalled by the Change Detector. Hence the average positions of both *A* and *B* must not be the same for two successive moments. If they were the same, no movement would be detected for either *A* or *B* for the second moment so that "ampliation of movement" would not take place and the phenomenological impression of two independent movements would be produced.

Let *P* be the time during which *A* and *B* are in contact. Then, if *P* contains two whole perceptual moments, the average position of both *A* and *B* will be the same for each of the two moments. Assuming that the phase of the moment train and the beginning of the period of contact are not correlated—a reasonable assumption, since there is no change in intensity of stimulation—then the chance of *P* containing two whole moments is

$$\begin{aligned} \frac{3D - P}{D} & \text{ when } 2D < P < 3D, \\ 1 & \text{ when } P > 3D, \text{ and} \\ 0 & \text{ otherwise.} \end{aligned}$$

If *P* contains only one whole moment (see Figure 3) then in each moment a movement of either *A* or *B* will be detected, so the ampliation condition is satisfied and perception of causality will occur. However, during one moment *A* and *B* will be perceived as being in contact so that the impression of 'stickiness' will arise. It may seem contradictory to use the position of *A* and *B* as part of the explanation of what happens in a moment when an illusory detection of movement has occurred. This objection also arises to Michotte's theory and he has considered it, quoting examples which show that: "The change in position of an object and the movement which brings about this change may co-exist as distinct data" (see p. 135). On the present theory, movement information is added to the input in the Change Detector so that higher processes have both movement and position information as input.

The chance of *P* containing one whole moment, so that the impression of 'sticky' launching arises, is

$$\begin{aligned} \frac{2D - P}{D} & \text{ if } D < P < 2D, \\ \frac{P - 2D}{D} & \text{ if } 2D < P < 3D, \text{ and} \\ 0 & \text{ otherwise.} \end{aligned}$$

If P contains no whole moments, then normal direct launching should be perceived. The chance that this occurs is

$$\begin{aligned} 1 & \quad \text{if } P < D, \\ \frac{P-D}{D} & \quad \text{if } D < P < 2D, \text{ and} \\ 0 & \quad \text{if } P > 2D. \end{aligned}$$

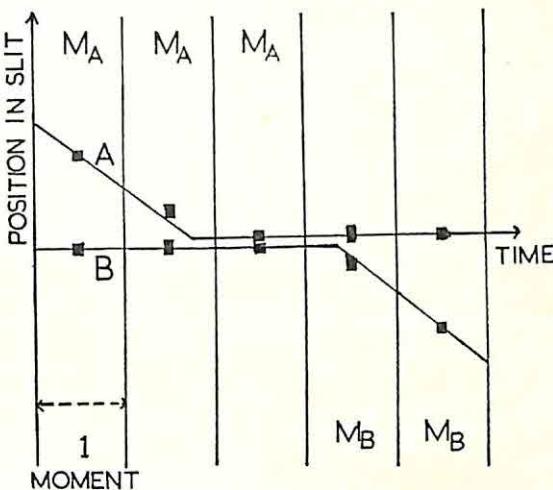


FIGURE 3. Perception of Causality according to the Perceptual Moment Model. The vertical lines are the boundaries of successive moments. The two other lines represent the physical positions of the objects, A and B , in the time about the period of contact. The small squares represent the average positions of the objects during the moments. In the moments labelled M_A , movement of object A is detected, and in those labelled M_B movement of B is detected. The case illustrated is an example of delayed launching since the average positions of the two objects are in contact for one moment and there is no break in the detection of movement.

If we compare these predictions with the results of Michotte's Experiment 29 the fit is extremely good (see Figure 4). Since the graph is the average of three observers, who will have different values of D , the flattened peak of the delayed launching curve is to be expected as are the tails of the other two curves.

The moving average model makes a qualitatively similar prediction to the perceptual moment model in that it predicts three stages: when the detection of movement is discontinuous; when the detection of movement is continuous, but A and B are perceived in contact; and when detection is continuous, but A and B are not perceived in contact. However the predicted time relationships are not those found empirically. If S is the time span of the moving average, and L is the time in which the standard is formed, the average position of A does not become constant until a time S after A has stopped, and even then,

movement of A will still be detected for a further time L . On the other hand, as soon as B moves, its average position will change and its movement will be detected. Therefore, we would expect direct launching if $0 < P < S$, delayed launching if $S < P < (S + L)$, and two independent movements if $P > (S + L)$. Now L is a variable which may not be constant, even in theory. Not enough has been worked out about the creation of the standard to make any firm predictions about this aspect of the models. However this is not crucial, since the boundary between direct and delayed launching does not depend on L but only on S , which should be constant in theory. In order to fit the moving average model it is necessary to assume that Michotte's three subjects have widely different and variable values of S . Hence the moving average model provides a much less satisfactory explanation than the perceptual moment model.

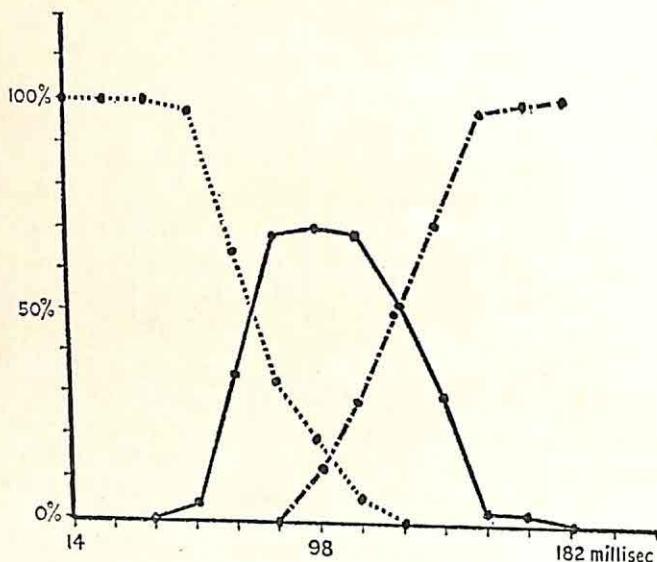


FIGURE 4. Average results for the three subjects of Michotte's Experiment 29 (from Michotte, A., 1963, *The Perception of Causality*; London: Methuen). Dotted line—Direct Launching; Continuous line—Delayed Launching; Dots and dashes—Two Movements.

A further corroboration of the perceptual moment model comes from another strange finding of Michotte's. He says: "It is worth noting that the observers very often mentioned that the Launching Effect was 'better' in cases where there was an interval of 30 to 40 msec at the point of impact than in cases where the second movement followed immediately. This must certainly be attributed to the length of time required for the 'rise of excitation', i.e., the length of time necessary for the observer to receive a fully developed sensory impression of Object A after it has halted at the centre. Indeed when the intervals are very short or non-existent, observers often have the impression that the two objects do not meet at all, there being neither contact nor impact".

The explanation that will be put forward here does not contradict Michotte's but is at a different conceptual level. On the perceptual moment model the movement of an object is detected both in the moment an object stops, and, paradoxically, in the following moment because its average position in the two moments is different. Also, when a movement starts, it is detected in that moment. Hence, if there is no gap between *A*'s movement and *B*'s, *B* will be perceived as moving, *before A* stops moving so that the impression of no meeting and no impact is natural. However, when *B* starts moving in the moment after *A* stops, this does not apply, and since all the other conditions for a good launching situation hold, the Launching Effect is seen as 'better'. On the moving average model, *A* appears to move for $(S+L)$ msec after it actually stops, whilst *B* appears to move almost immediately after it starts. (As in all this explanation of the perception of causality, the threshold for movement detection has been ignored. Introducing it only complicates the theory, but does not lead to any significant alterations in the conclusions that can be drawn.) Consequently in all cases where launching is perceived, *B* starts moving before *A* stops, and in all the Direct Launching cases there should appear to be no contact. So the perceptual moment model is even more strongly supported.

Michotte's numerous other experiments could also be discussed, but to deal with them in as much detail as we have dealt with the present experiment would lead to a very lengthy paper. The main elaboration necessary would be a more precise specification of the threshold for the detection of change. This would be particularly relevant to the experiments in which object *A* and object *B* moved at different speeds. Superficially, at least, there does not seem to be anything in the other experiments to alter the conclusion of the present analysis, that the perceptual moment model is much the most plausible model, since the G.M.A. model would provide an explanation that was very similar to and as inadequate as the moving average model.

V. DISCUSSION

The one other situation, known to the writer, which satisfies the condition of no large change in the intensity of the stimulus is Ansbacher's (1944) work on the Brown Circle Illusion. However, all the models can explain the phenomenon. This leaves the perceptual moment model as superior to the moving average model empirically and theoretically, and far superior to the G.M.A. model empirically. Two further issues will now be raised. First, a closer examination will be made of some of the assumptions common to the three models. Then consideration will be given to the way in which the perceptual moment model can be adapted to deal with situations in which a sharp change in stimulus intensity occurs.

(i) *The common assumptions* The first assumption of the present thesis is that the basic units for the detection of change are molar, not molecular, that is an accumulator model is preferred to one based on pulse interval

modulation (P.I.M.). Apart from the functional argument that a P.I.M. model would impose a much greater computational load, especially with a standard changing constantly due to motor movements, Barlow (1961) points out that there is physiological evidence (Fitzhugh, 1957) that what matters is the aggregate number of impulses in a time interval which is long in comparison with an interpulse interval. The only case known to the writer where P.I.M. has been used in an explanatory model is by Kelly (1961) for flicker fusion (C.F.F.). The writer has recently shown that this model contains logical inconsistencies, which can only be removed by dispensing with the use of P.I.M. (Shallice, 1964).

The second assumption is that a spatial averaging process of sorts takes place in the accumulator. If no spatial averaging took place, a moving object would be represented in the accumulator by the distribution of all the positions it took during that moment, and more importantly a stationary object would also be represented by such a distribution if the eyes were moving during the moment. In order to detect that the stationary object had not moved, a standard would have had to be created which mirrored that distribution. Obviously to create the proper distribution for the standard is a much more difficult task than to create simply the average position where an object should be found. Again we use the idea that introducing a spatial averaging process will greatly reduce the computational load later in the total system. The physiological counterpart to averaging is probably lateral inhibition, and the best psychological evidence comes from the study of apparent movement and Ansbacher's (1944) work on the Brown Circle Illusion.

One part of the model which has remained relatively undeveloped is how the standard is created. The reason for this is that it did not seem of vital importance when comparing the models. Essentially, this is a very similar problem to the old problem of the Stability of the Visual World, since the standard will correspond to what Gregory (1958) calls the signals from the eye/head system, while the input will correspond to the signals from the image/retina system. Gregory put forward his cancellation theory, an extension of Helmholtz's, as an alternative to MacKay's (1958) theory. Yet the present model also seems compatible with MacKay's theory; in fact it has close similarities, as one can judge from MacKay's summary of his theory: "If perception is the adaptive 'keeping up to date' of an organism's state of organization for activity, then what requires informational justification is not the maintenance of stability, but the perception of change." Maybe the two theories differ mainly in their frames of reference. While the theories of the Stability of the Visual World provide a way of attacking the problem of the creation of the standard, to deal with it here it would be necessary to consider a whole set of experiments which are only vaguely related to the perceptual moment theory. So that apart from this hint of a method, this part of the model will be left almost as undeveloped as at the beginning of this paragraph.

(ii) *The effect of quick large changes in intensity* In the perceptual moment model the accumulator has to discharge at specific times. Hence there must be

some mechanism which acts as the 'neural clock' and triggers off the discharge of the accumulator. In the present model it will not have to behave like an ordinary clock, as the speed at which it runs will vary with stimulus intensity.

A speculation about this mechanism will now be put forward, the main justification for which is that it is very simple and yet agrees with much of the evidence. The proposed mechanism consists solely of an accumulator. The accumulators that have been discussed before have all been fed by only a small number of receptors, maybe only one, and will be called specific accumulators.

Consider a general accumulator, which receives input from all sense receptors, and assume that as soon as its contents reach a fixed level it will discharge and the discharge can act as a timing signal which will trigger the discharges of

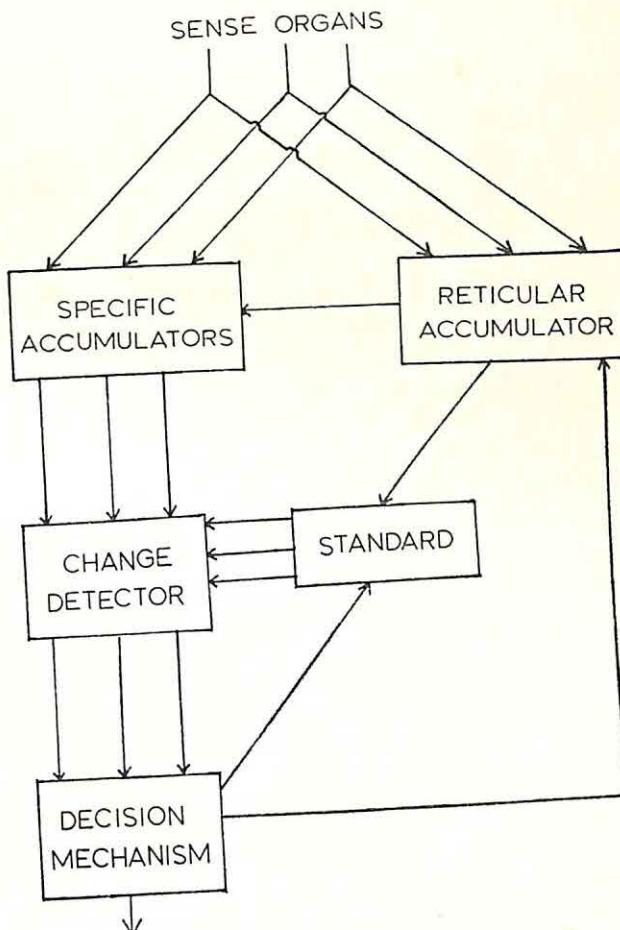


FIGURE 5. Elaboration of the Perceptual Moment Model. The Decision Mechanism represents higher processes not dealt with in this paper. Its output to the Reticular Accumulator is to allow for the partial cortical control of arousal which is known to exist. The outputs of the Reticular Accumulator are the timing pulses which trigger the discharge of the Specific Accumulators and the Standard.

the specific accumulators. After the discharge it will start accumulating again and the cycle will repeat indefinitely. Thus, we have a very simple mechanism which produces a discharge at regular intervals under steady stimulus conditions, but if the stimulus intensity is increased the rate of discharging increases. This increase would not be as great as the increase in the rate of firing of the nerve fibres upon which the stimulus acts, since the vast majority of the input channels into the general accumulator will be unchanged. The general accumulator will be known as the reticular accumulator, since the obvious place for such a mechanism is in the reticular system where impulses from all sensory pathways converge (see Figure 5).

This suggestion is supported by Lindsley's (1958) finding that the two-flash threshold is reduced by stimulation of the reticular system. Since the E.E.G. is thought to be controlled by the reticular system, the speculation is also supported by the known correlation between the period of the E.E.G. and the perception of change (see Murphree, 1954; Surwillo, 1963). This was originally predicted from a theory, obviously related to the present one, that the alpha rhythm was acting as a scanner of the visual input (see Walter, 1950; Wiener, 1948), but the movement illusions that one would predict from the theory are not found (see MacKay, 1953).

If the reticular discharges and the E.E.G. phase are related then the correlation between reaction time and alpha rhythm phase seems much smaller than would be predicted. This was known to be a difficulty for the visual scanning theory as well. The point at issue is that if a signal arrives just as the moment is about to close, it should be delayed in the accumulators for a much shorter time than a signal arriving at the beginning of the moment. In fact, a difference of only about 6 msec between the effects of stimulation at the most and least favoured phases was found by Lansing (1957) and Callaway (1962), and Walsh (1952) found no correlation with phase at all.

However, if the light is relatively weak, the consequent R.T. is usually long and the signal would probably not be detected in the first moment. Thus which phase will be the favoured one will be unclear, since the answer depends upon the chance of a signal being detected in the moment in which the signal arrived, relative to that obtaining when it is present for the whole moment. Consequently the resulting variability in R.T. will ensure that the effect of phase is much less than previously expected. When the light is strong, then whether the stimulus is detected in the moment in which it arrives is still a relevant although less important issue, but there is a further complicating factor. If the stimulus arrives early in the moment, then the moment will close earlier than had been anticipated, since the reticular accumulator will fill more quickly than before and the timing signal for the closing of the moment will arrive earlier at the specific accumulators. Therefore the difference in R.T. between the most and least favoured phases for detection will be reduced for two reasons.

Since strong stimuli, especially visual, affect the rate at which the reticular accumulator fills, bright flickering lights will produce a series of moments

varying in length. This will contribute to the Bartley Brightness Enhancement Effect, since, at certain speeds, the flashes will tend to fall into the longer moments, so that the contrast between the output of the accumulator in these moments and those preceding and following will be greater than if all the moments had the same length. In addition, since the moments vary in length and the 'standards' with which current stimulation is compared are produced on the assumption of a constant situation, movement effects will be perceived. They will arise because the effects of eye movements are not correctly cancelled, as the period over which spatial averaging takes place is different from the one which provided the basis of the standard. This is a possible explanation for the well-known movement illusions produced by bright flickering light.

This hypothetical mechanism will have to stand up to much more severe tests before it is more than a speculation, but it does show one way in which a variable moment length could be made an intrinsic part of the model. Two areas in which it might be tested are those of flicker phenomena and visual masking. One difficulty with both these areas is their complexity, and where visual masking is concerned there is the added difficulty that some of the experiments will require an explanation which deals with recognition as well as the detection of change. Recognition must be mediated by higher level mechanisms, not subsumed in the present variety of hypothesis.

VI. APPENDIX

To show that $N = TJ/(1 - e^{-bT})$ is not a good approximation to the empirical (N, T) curve consider:

$$\begin{aligned} h(T) &= \frac{N(2T)}{N(T)} \\ &= \frac{2TJ}{1 - e^{-2bT}} \cdot \frac{1 - e^{-bT}}{TJ} \\ &= \frac{2}{1 + e^{-bT}}. \end{aligned}$$

Now if $h(T) = 1 + v$ where $v \ll 1$,

$$\text{then } \frac{2}{1 + e^{-bT}} = 1 + v.$$

$$\therefore e^{-bT} = 2[1 - v + O(v^2)] - 1$$

$$= 1 - 2v, \text{ to the first order in } v.$$

$$\therefore e^{-2bT} = 1 - 4v, \text{ to the first order in } v.$$

$$\therefore h(2T) = \frac{2}{1 + e^{-2bT}} = 1 + 2v, \text{ to the first order in } v.$$

Put $T = 50$ msec, then $2T = 100$ msec, and the Bunsen-Roscoe Law holds for absolute thresholds and small areas, at least to a high degree of approximation, so that $N(2T) = N(T)(1+w)$, where w is very small.

∴ If the G.M.A. is a satisfactory model we should have

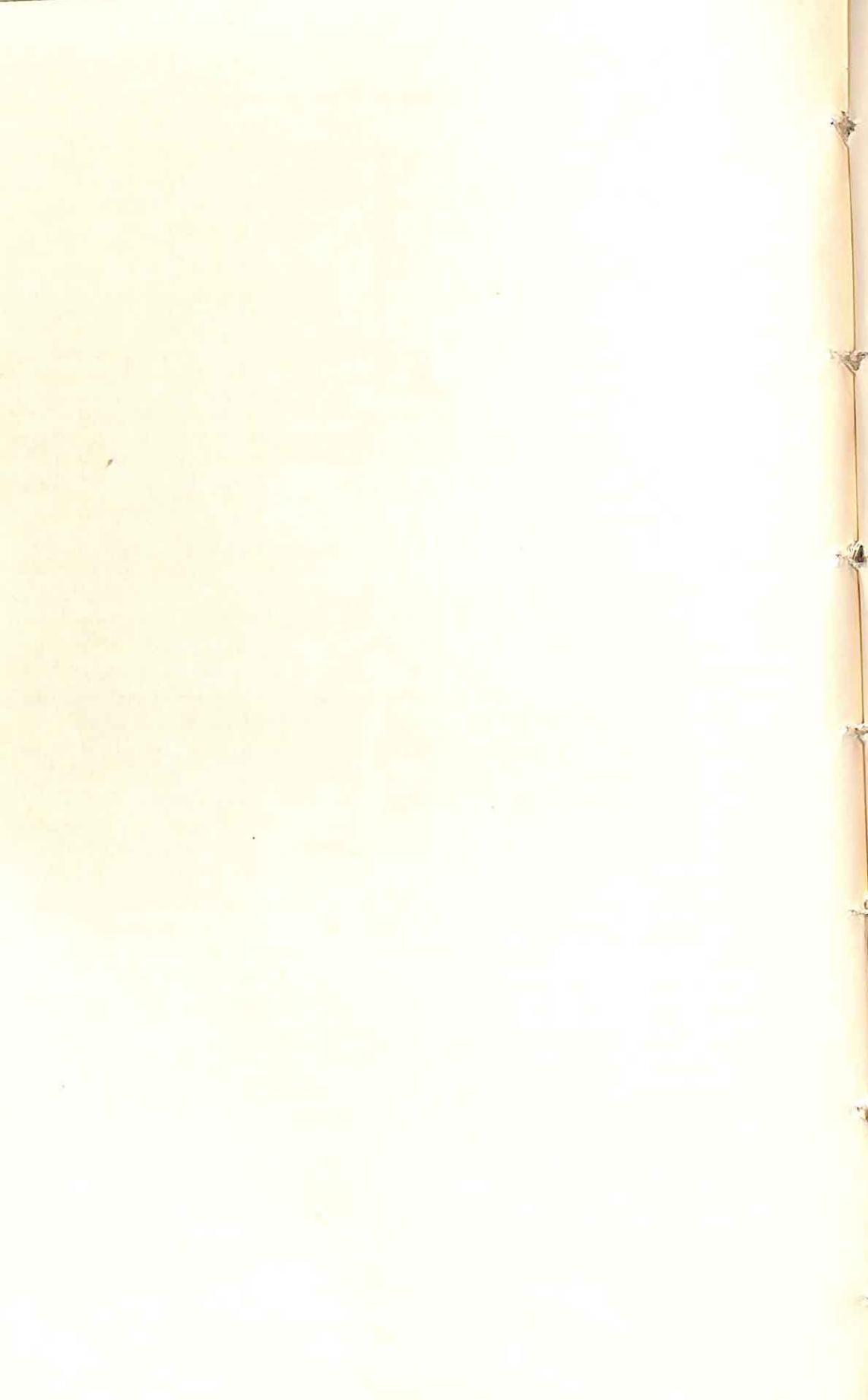
$$N(200 \text{ msec}) = N(100 \text{ msec})(1+z), \text{ where } z=2w.$$

However, for high values of Q , $z \approx 1$ (Graham and Kemp, 1938), and for low values of Q , $z \approx 0.5$ (Bouman and van der Velden, 1947), so that in both cases $z \neq 2w$.

REFERENCES

- ANSBACHER, H. L. (1944). Distortion in the perception of real movement. *J. exp. Psychol.*, **34**, 1-23.
- AUDLEY, R. J. (1963). Decision models in reaction time. Paper read at the International Congress of Psychology, Washington, 1963.
- BARLOW, H. B. (1961). Possible principles underlying the transformations of sensory messages. In *Sensory Communication* (ed. W. A. Rosenblith). New York: M.I.T. and John Wiley.
- BEKESY, G. von (1936). Low-frequency thresholds for hearing and feeling. *Ann. Physik*, **26**, 554-566. Reprinted in *Experiments in Hearing* (ed. E. G. Wever), 1960. New York: McGraw-Hill.
- BLACKWELL, H. R. (1963). Neural theories of simple visual discriminations. *J. opt. Soc. Amer.*, **53**, 129-160.
- BOUMAN, M. A. (1961). History and present status of the Quantum theory in vision. In *Sensory Communication* (ed. W. A. Rosenblith). New York: M.I.T. and John Wiley.
- BOUMAN, M. A., and VAN DER VELDEN, H. A. (1947). The two quantum exploration of the dependence of the threshold values and visual acuity on the visual angle and the time of observation. *J. opt. Soc. Amer.*, **37**, 908-919.
- BOYNTON, R. B., and SIEGFRIED, J. B. (1962). Psychophysical estimates of on-responses to brief light flashes. *J. opt. Soc. Amer.*, **52**, 720-721.
- BRINDLEY, G. S. (1960 a). Two more visual theorems. *Quart. J. exp. Psychol.*, **12**, 110-112.
- BRINDLEY, G. S. (1960 b). *Physiology of the retina and the visual pathways*. London: Edward Arnold.
- CALLAWAY, E. III (1962). Factors influencing the relationship between alpha activity and visual reaction time. *E.E.G. Clin. Neurophysiol.*, **14**, 674-682.
- DAVIS, H. (1951). Psychophysiology of hearing and deafness. In *Handbook of Experimental Psychology* (ed. S. S. Stevens). New York: John Wiley.
- DREW, G. C. (1937). The variation of sensory thresholds with the rate of application of the stimulus. (Part III.) *Brit. J. Psychol.*, **27**, 297-302.
- FITZHUGH, R. (1957). The statistical detection of threshold signals in the retina. *J. gen. Physiol.*, **40**, 925-948.
- GRAHAM, C. H., and KEMP, E. H. (1938). Brightness discrimination as a function of the duration of the increment in intensity. *J. gen. Physiol.*, **21**, 635-650.
- GRANT, E. L. (1952). *Statistical Quality Control*. New York: McGraw-Hill.
- GREGORY, R. L. (1958). Eye movements and the stability of the visual world. *Nature, Lond.*, **182**, 1214-1216.
- GRINDLEY, G. C. (1936). The variation of sensory thresholds with the rate of application of the stimulus. (Parts I and II.) *Brit. J. Psychol.*, **27**, 86-95 and 189-195.

- JOHNSON, E. P. (1958). The character of the *b*-wave in the human electroretinogram. *Archiv. Ophthal.*, **60**, 565-589.
- KELLY, D. H. (1961). Visual responses to time-dependent stimuli. II. Single-channel model of the photopic visual system. *J. opt. Soc. Amer.*, **51**, 747-754.
- LAMING, D. R. J. (1962). A statistical test of a prediction from information theory in a card-sorting situation. *Quart. J. exp. Psychol.*, **14**, 38-48.
- LANSING, R. W. (1957). Relation of brain and tremor rhythms to visual reaction time. *E.E.G. Clin. Neurophysiol.*, **9**, 497-504.
- LICKLIDER, J. C. R. (1951). Basic correlates of the auditory stimulus. In *Handbook of Experimental Psychology* (ed. S. S. Stevens). New York: John Wiley.
- LINDSLEY, D. B. (1958). The reticular system and perceptual discrimination. In *Reticular Formation of the Brain* (ed. H. H. Jasper et al.). Boston: Little, Brown.
- MACKAY, D. M. (1953). Some experiments on the perception of patterns modulated at the alpha frequency. *E.E.G. Clin. Neurophysiol.*, **5**, 559-562.
- MACKAY, D. M. (1958). Perceptual stability of a stroboscopically-lit visual field. *Nature, Lond.*, **181**, 507-508.
- MACKAY, D. M., and McCULLOCH, W. S. (1952). The limiting information capacity of a neuronal link. *Bull. Math. Biophysics.*, **14**, 127-135.
- MICHOTTE, A. (1963). *The Perception of Causality*. London: Methuen.
- MUELLER, C. G. (1954). A quantitative theory of visual excitation for the single photoreceptor. *Proc. Nat. Acad. Sci.*, **40**, 853-863.
- MURPHREE, O. D. (1954). Maximum rates of form perception and the alpha rhythm: an investigation and test of current nerve net theory. *J. exp. Psychol.*, **48**, 57-61.
- PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100-115.
- ROBERTS, S. W. (1959). Control chart tests. *Technometrics*, **1**, 239-250.
- SHALLICE, T. (1964). Kelly's model of the photopic visual system (to be published).
- STONE, M. (1960). Models for choice-reaction time. *Psychometrika*, **25**, 251-260.
- STROUD, J. M. (1949). The psychological moment in perception. In *Trans. Sixth Conference on Cybernetics* (ed. Van Foerster). New York: Josiah Macy, Jr.
- STROUD, J. M. (1955). The fine structure of psychological time. In *Information Theory in Psychology* (ed. H. Quastler). Glencoe, Ill.: Free Press.
- SURWILLO, W. W. (1963). The relation of simple response time to brain-wave frequency and to age. *E.E.G. Clin. Neurophysiol.*, **15**, 105-114.
- WALSH, E. G. (1952). Visual reaction time and the alpha rhythm, an investigation of the scanning hypothesis. *J. Physiol.*, **118**, 500-508.
- WALTER, W. G. (1950). The twenty-fourth Maudsley lecture: The functions of electrical rhythms in the brain. *J. ment. Sci.*, **96**, 1-31.
- WEVER, E. G., BRAY, C. W., and WILLEY, C. F. (1937). The response of the cochlea to tones of low frequency. *J. exp. Psychol.*, **20**, 336-349.
- WHITE, C. T. (1963). Temporal numerosity and the psychological unit in duration. *Psychol. Monog.*, **77**, No. 12.
- WIENER, N. (1948). *Cybernetics*. New York: Wiley.



FITTING A LINEAR TRACE DECAY MODEL TO TASTE
COMPARISONS AND PREFERENCES¹

By R. A. M. GREGSON

University of Canterbury, New Zealand

Taste assessment procedures are examples of human sensory and hedonic judgements which have not generally been analysed in dynamic psychophysical terms, though an extensive statistical literature has grown up to cover the use of such procedures. A taste comparison procedure, using mixtures of wines as stimuli and 168 non-expert subjects as tasters, yielded data on five response measures which have been analysed using a linear trace decay model. The fit of the model is good and accounts for most of the observed variation in assessed differences which is associated with orders of tasting samples. Reasons for a poor fit in a minority of cases are discussed.

I. INTRODUCTION

This paper describes an examination of data from a taste assessment procedure, using a linear trace decay model to account for the detailed pattern of results obtained. Whilst the linear trace decay idea is familiar to most experimental psychologists, if only because it underlies some explanations of short term memory or of the negative time error or time-order effect, taste assessment procedures are not generally known to psychologists and some general comment on their form and use, and their attendant psychophysical problems, is therefore pertinent.

The food industry has developed a number of procedures for describing or evaluating tastes and other properties of food, which are essentially examples of applied psychophysics. Such procedures have in general been developed by trial and error, more recently by statisticians consulting to the food industry [8, 10, 12, 35], and usually without much overt regard to psychophysical theory or precedents, although the psychometric problems are logically similar as Thurstone showed [31]. They are called 'assessment' procedures because the comparisons involved are usually between a standard which is acceptable on both economic and psychological grounds and a slight modification of the standard whose taste acceptability is to be tested against the standard.

The procedures used by food tasters involve rating, matching, identifying or discriminating between tastes, using—as appropriate—single or multiple taste stimuli, and verbal or numerical classifying, rating or scoring systems. Collateral chemical information on the stimuli is often available, and as might be expected the stimulus error is often committed [32]. Parallels to food assessment procedures can readily be identified in the literature of attitude scale

¹ I would like to express gratitude to Mr. P. M. Shaw for critically reviewing early drafts of this paper, and to the Directors of J. Lyons and Company Limited, London, in whose tasting laboratories the experiment reported was done.

construction, multidimensional scaling, attribute classification or classical psychophysics. Statistical models advanced to serve as quantifying paradigms for food assessment methods often have an oddly static quality [18, 17, 27] as though time errors, adaptation level effects, and serial interference phenomena never existed or could be ignored.

The literature of taste methodology frequently advocates procedures which are identifying or matching tasks: they stem in the main from Helm and Trolle's work [21, 35] on the triangle test, or from what is called the duo-trio test. Respectively these are:

The triangle test. Present the subject with three samples XXY (position of Y suitably randomized over successive trials) and ask him to pick out the odd one. It is assumed that the probability of doing this by chance is $1/3$, irrespective of the serial order of tasting, whether the samples vary on one or many subjective dimensions, and whatever the subject's prior expectations are.

The duo-trio test. Present the subject with X and Y , and then, after he has tasted them, present a third sample (X or Y) and ask for this to be matched to one of the first two. It is assumed that the probability of doing this by chance is $1/2$, with the same assumptions as for the triangle test. The statistical power of various tests has been calculated on these assumptions [18] and a number of attempts have been made [2, 3, 12, 13] to take account of individual differences by complicating the statistics in various ways. All such models assume a constant single probability, for a given subject, of correctly performing the task. In attempts to increase the power of such procedures, the choice of say, the odd two samples out of five or the odd three out of eight has been advocated. One can only boggle at the psychophysical naïvety of such suggestions in which the human observer is so readily treated as a static bioassay.

Consider a case where we had, say, a mixed population of judges, some of whom were known to be completely blind and others completely sighted, and to whom we presented three picture postcards identical in size and texture cues but two bearing the same reproduction of a Rembrandt and the third a reproduction of a Vermeer. Granted the independence of discriminatory and hedonic judgements, the plausibility of the statistical assumptions underlying the *triangle* or *duo-trio* procedures can be seen; they are in fact advocated precisely because their attendant statistical analysis is supposedly simple and legitimate; in short, a single parameter binomial probability model. This type of assumption, in a slightly more sophisticated form, underlies the *method of complete triads* where we present stimuli X and Y , then Z , and ask the subject to judge whether Z is more similar to X or Y . This method is used in multi-dimensional scaling [33], and is very similar to the *t-s-r* (triad-similarity-rating) procedure used here.

Empirical studies conducted to see precisely what happens in a taste assessment procedure—its 'internal psychophysics', if we may be permitted the term—do exist, but are few [22, 28]. Attempts to see if the internal psychophysics are such as to cast doubt on the validity of the choice of statistical axioms used for significance testing are probably even fewer [16, 17]. Discussion

of the measurement problems involved, and the methods used [5, 6, 20, 27] exist, but we feel there is point in trying to fit a dynamic rather than static model to taste comparison and preference data, for three reasons:

- (i) To see whether a statistical or formal analysis of such data can usefully be made on results not specially gathered for the purpose of such analysis, because our analytical procedures are more satisfactory if they can be used in this way rather than being restricted to 'contrived' situations. In this sense we want to perform what some statisticians would call a retrieval analysis.
- (ii) To see whether a model which relates the 'internal psychophysics' of a procedure to the general body of psychophysical results in taste and in other sensory modalities can be fitted to data, because there is value in demonstrating the generality of a theory in both fundamental and applied fields.
- (iii) To see what implications the analysis has for considerations of predictive validity, within the field of taste comparisons.

For these reasons, there would seem to be some point in taking an existing taste assessment procedure and examining it as a judgement process without specially modifying it. We have, therefore, chosen a procedure from an unpublished comparative examination of procedures which varied in their form of instructions or in the order in which discrimination and preference judgements were elicited, which is not trivially simple, for which sufficient data exist, and which we feel illustrates some of the features and problems of statistical and psychophysical method in the context of taste measurement.

II. EXPERIMENT

Stimuli

The original stimuli used were mixtures of two wines, a Graves (undated) and a 1957 Barsac, both in bottle and in one batch from the same cellars. Three pairs of mixtures were actually tasted, as follows. In the 9/1 experiment:

$$\begin{aligned} A &= \text{a mixture of 90 per cent Graves and 10 per cent Barsac} \\ B &= \text{a mixture of 10 per cent Graves and 90 per cent Barsac,} \end{aligned}$$

in the 8/2 experiment

$$\begin{aligned} A &= \text{a mixture of 80 per cent Graves and 20 per cent Barsac} \\ B &= \text{a mixture of 20 per cent Graves and 80 per cent Barsac,} \end{aligned}$$

in the 7/3 experiment

$$\begin{aligned} A &= \text{a mixture of 70 per cent Graves and 30 per cent Barsac} \\ B &= \text{a mixture of 30 per cent Graves and 70 per cent Barsac.} \end{aligned}$$

In each case the *B* mixture is predominantly Barsac. Mixtures were employed rather than a straight comparison because it was desired to vary the difference in degree; the difference in the character of the two wines is lost provided the concentration of any taste-bearing substance in one wine does not fall below threshold in the more extreme (9/1 or 8/2) mixtures. The mixtures reduce the sweetness difference to near-threshold, and the colour difference was lost under the viewing condition used (yellow light, wine in glasses set on yellow paper). The Barsac was sweeter, more yellow, and had what professional

tasters call more 'body': overall it is taken to be the stronger stimulus, in the sense that *B then A* is a more difficult pair to discriminate than *A then B*, provided time errors are negative. Fresh bottles were opened for each day's tasting, the mixtures tasted in wine glasses of about 20–30 c.c. portions, the mixing being done at the time of decanting.

Subjects

In all 168 subjects participated, 48 for the 9/1 mixtures, 72 for the 8/2, and 48 for the 7/3. The subjects were factory and office staff of both sexes and a wide age range.

Procedure

The subject is presented with three samples, of which two are identical and the third is different. We will call the two taste substances *A* and *B*, so that in a set of three—here called a triad—we have, reading from the subject's left to right, the following spatial orders;

Order I	:	<i>A</i> ₁ <i>A</i> ₂ <i>B</i>
Order II	:	<i>A</i> <i>B</i> ₁ <i>B</i> ₂
Order III	:	<i>B</i> ₁ <i>B</i> ₂ <i>A</i>
Order IV	:	<i>B</i> <i>A</i> ₁ <i>A</i> ₂

where either *A*₁ and *A*₂ or *B*₁ and *B*₂ are identical. Consider the presentation order I: we present *A*₁ *A*₂ *B* with *A*₂ as the middle and last tasted stimulus, the order of tasting being: left, right, middle. The subject expresses a preference between *A*₁ and *A*₂, by identifying the preferred stimulus and rating his preference on a five point scale; this is called the 'identical paired comparison'. He then does the same for *A*₂ and *B*; this is called the 'different paired comparison'. Then *A*₁ and *A*₂ are spatially interchanged, the whole new triad *A*₂ *A*₁ *B* is retasted in the order left, right, middle, and the middle stimulus is rated on a five point scale for its relative similarity to *A*₂ and *B*. All the presentation orders I, II, III and IV occur equally often in a total experiment. In the experiment here reported, each subject did three trials in succession, the design being such that any one subject did three of the possible four orders, and each order occurred equally often as 1st, 2nd or 3rd triad encountered.

Each subject tasted the wines in isolation from other subjects, in tasting cubicles, only the experimenter being present. The wines were served at room temperature (warmer than is usual for table wines of these types) and palate cleansers in the form of rye crisp-bread biscuits were eaten before each triad of samples, but not between samples.

Instructions

The subjects were told that the experiment was concerned with tasting white table wines, and the use and importance of the palate cleansers emphasized. They were then told:

"Here are three samples of wine (experimenter sets the glasses out spaced evenly), would you please taste first the right-hand sample, then the left-hand sample, and finally the middle sample. Take small sips. Would you now say how much you prefer or don't prefer the middle one or the right-hand one, using this card". (A card was presented to S by E bearing the statements: "Prefer this one very strongly." "Prefer this one strongly." "Prefer this one." "Prefer this one slightly." "No real preference for either". These five statements were given Likert values 5 to 1 respectively in the treatment of results, although no numerical values were ever shown to subjects.)

This procedure was repeated with the middle and left-hand glasses. The two actually identical mixtures were then interchanged, and the subject asked :

"Taste them again, first the right, then the left and finally the middle one, and say how the middle one compares with the other two, using this card." (Card presented

bearing the statements: "Same as the left-hand one." "More like the left-hand one." "Equally different from both right and left." "More like the right-hand one." "Same as the right-hand one"). This procedure we have called *t-s-r* (triad-similarity-rating) and Likert values 1 to 5 are given to the five statements but in such a manner that the correct statement is always scored 5 irrespective of its laterality.

III. THEORY

A linear trace decay model [7, 23] will be used to make predictions about the effects of each of the four presentation orders I, II, III, IV.

To construct any theory to account for the pattern of perceived and valued taste intensities in our experiment, we need a notation, and some assumptions which we make implicit or embodied in the notation. In the model we treat all subjects as replicates of each other, and assume they all behave in the same way as regards their self-paced tastings. Let W (standing for A or B) be the effective intensity of a stimulus at a point in time. We consider the three stimuli in a triad to be tasted at points in time which we will call t_0 , t_1 and t_2 successively. Now for a linear trace decay the value of W falls with time. With the assumption that $t_1 - t_0 = t_2 - t_1 = k$ (a constant interval), W will decrease over the interval by a multiple of W , say δW , where $1 \geq \delta \geq 0$.

In practice the intervals between tastings are uncontrollable to very precise limits, but timing a sample of tastings showed the interstimulus interval to be very regular and modal at about six seconds: the actual time intervals are all longer than the short interval in which positive time errors have been reported [26]. Changes in the direction of negative time errors have more usually been reported in other sensory modalities.

Let A , B also be used to refer to the two stimulus intensity levels in a given triad, and further assume $B = (1 + \pi)A$, where $1 \geq \pi \geq 0$. The linear trace decay of an A stimulus will then be δA , but of a B stimulus will be $\delta B = \delta A + \delta \pi A$. The term $\delta \pi A$ is assumed to be second-order and will therefore be neglected.

We distinguish between the effective stimulus intensities—if we like to speculate these could be what have been called the 'proximal stimuli' or neural correlates of peripheral stimulation—and the subject's response, because the subject expresses the effective stimulus intensity, or some comparison between such intensities, by a response choice which is converted in turn to a probability or a rating on a scale.

The actual response measure we use presupposes some psychophysical transformation ψ within the subject, hence we may write $\psi(W)$ for the subjective estimation of a stimulus magnitude, and $h\{\psi(W)\}$ for a hedonic evaluation: here the notation is intended to imply that $\psi(W)$ is a necessary prior step to $h\{\psi(W)\}$. We will assume, so as to simplify the argument, that $\psi(W)$ is linear on W , $h\{\psi(W)\}$ is linear on $\psi(W)$, and recorded measures ψ' or h' of ψ or h responses respectively are linear on $\psi(W)$ or $h\{\psi(W)\}$ over the stimulus magnitude range considered.

In this sense ψ and h are intervening variables, ψ' and h' are response indicators [15] of these variables, and the subsequent argument is to be taken as interpretable in ψ' , h' or ψ , h measures. For convenience, therefore, we will write $\psi(W)$ and $h(W)$ as an abbreviated notation for $\psi'(W)$ or $\psi(W)$ and $h'(\psi(W))$ respectively throughout; the meaning in functional notation should be clear from the context.

Considering presentation order I , we have

$$\begin{aligned} &A_1 \text{ at } t_0, \text{ therefore } A - 2\delta A \text{ at } t_2; \\ &B \text{ at } t_1, \text{ therefore } A + \pi A - \delta A \text{ at } t_2; \\ &A_2 \text{ at } t_2, \text{ therefore } A \text{ at } t_2; \\ &\quad \text{for } t_1 - t_0 = t_2 - t_1. \end{aligned}$$

The identical paired comparison is A_1 with A_2 or $(A - 2\delta A)$ with A at t_2 . The different paired comparison is B with A_2

$$\begin{aligned} &\text{or } (1 + \pi - \delta)A \text{ with } A \text{ at } t_2, \\ &\text{or } [1 + \pi - (k+1)\delta]A \text{ with } (1 - k\delta)A \text{ at } t_{k+2}, \end{aligned}$$

allowing for delay between tasting and evaluation, where k is any multiplier to stretch the time interval for which δA is the amount of trace decay in unit time. The t - s - r judgement is

$$\begin{aligned} &A - 2\delta A \text{ with } A \text{ at } t_2; \\ &A + \pi A - \delta A \text{ with } A \text{ at } t_2. \end{aligned}$$

Hence the judgements elicited are:

$$\text{For the identical pair, } h(2\delta A); \quad (1)$$

$$\text{for the different pair, } h[(\pi - \delta)A]; \quad (2)$$

$$\text{for the } t\text{-}s\text{-}r, \psi[(\pi - \delta)A] > \psi(2\delta A). \quad (3)$$

Given the linearity of $\psi(W)$ on W ,

$$\pi > 3\delta \text{ for correct matching.}$$

If only the largest difference controls the response, i.e., if $|\pi - \delta| > |2\delta|$ with $\pi \gg \delta$ then for the t - s - r the judgement elicited is

$$\psi[(\pi - \delta)A], \quad (4)$$

and the recorded measures are linear functions of (1), (2) and (4).

Therefore, for each of the four presentation orders we have responses with the theoretical relative magnitudes given in Table I.

TABLE I. THEORETICAL RESPONSE MAGNITUDES FOR EACH TASTING ORDER AND RESPONSE MEASURE.

Tasting order	Identical pair	Different pair	t - s - r
I (ABA)	$h(2\delta A)$	$h[(\pi - \delta)A]$	$\psi[(\pi - \delta)A]$
II (ABB)	$h(\delta A)$	$h[(\pi + 2\delta)A]$	$\psi[(\pi + 2\delta)A]$
III (BAB)	$h(2\delta A)$	$h[(\pi + \delta)A]$	$\psi[(\pi + \delta)A]$
IV (BAA)	$h(\delta A)$	$h[(\pi - 2\delta)A]$	$\psi[(\pi - 2\delta)A]$

Response measures

For any given response measure employed over all the 12 terms of the above table, for the three stimulus mixtures 9/1, 8/2 and 7/3, we can obtain estimates of their relative magnitudes; with the restriction that we cannot, by definition, use the same metric for h and ψ .

From the choice or rating response which each subject made during the experiment, as described above, we have derived five response measures; four of them are hedonic and symbolized as h_1 to h_4 , the fifth based on the *t-s-r* ratings is a measure of similarity symbolized as ψ_5 . It is stressed that there is an implicit ψ function in all five measures; hence the enumeration of suffixes. For the rating scale measures h_2 , h_4 and ψ_5 the zero or scale unit is in part arbitrary; we have chosen numbers to give comparable magnitudes throughout, but similarities between the obtained ($\hat{\pi}$ and $\hat{\delta}$) estimates for different response measures should be regarded with reserve.

Identical pair proportionate preference, h_1 : in each triad tasted, the 'identical pair' preference (between A_1 and A_2 or between B_1 and B_2) was recorded as a straight choice, with 'equals' judgements allowed and scored ($\frac{1}{2}$, $\frac{1}{2}$) for conversion to proportions; straight preferences scored (1, 0) or (0, 1). All the preferences for triads of one order were pooled, and the proportions of all such triads in which the second stimulus (A_2 or B_2) was preferred was obtained. Each triad order is different in trace decay terms (though not in physical terms) and must therefore be considered separately. The pooled proportionate preferences for each triad order under each mixture condition are converted to unit normal deviates by treating each proportionate preference as an area under the cumulative normal curve [33].

Identical pair pooled ratings preference, h_2 : corresponding to h_1 above, but in rating form instead of choice form, the strength of preference was recorded for each identical comparison ($1 \leq$ preference strength ≤ 5), the pooled frequency distribution of such ratings over all conditions was obtained, and its mean and *s.d.* used to express the mean rating for each of the 12 triad-order and stimulus mixture conditions in standardized deviate form. Negative values here do not, then, indicate reversal of preference but indicate less than average preference strength, with direction unspecified.

Different pair proportionate preference; h_3 : this measure is obtained in exactly the same manner as h_1 , but is based on pooled proportionate preferences for the different pairs (between A and B) in each triad; transformed in the same way and using the same scoring conventions.

Different pair pooled ratings preference; h_4 : as for h_2 we have ratings of the degree of preferences for one or other member of the different pair, but here we have taken account of the direction of preference and not merely the strength, and combined two 5 point scales back-to-back to make one 9 point bipolar scale. This can be done because rating '1' = 'indifference' on each 5 point scale (one scale for each direction of preference, A preferred to B or B preferred to A) becomes the mid-category of the 9 point scale running from 'strong preference

for A' to 'strong preference for B ' ($1 \leq \text{preference} \leq 9$). These combined ratings were pooled to obtain the $s.d.$ of the rating total frequency distribution, but in this case the standardized values are expressed about the mid-point of the frequency distribution, not about their pooled frequency distribution mean. This implies that each standardized value has a constant added to it, but gives an obvious meaning in terms of preferences to the sign of each standardized deviate value. The scales combined back to back each yielded a centrally tending frequency distribution skewed towards 'indifference', but the frequency of 'indifference' responses was sufficiently large for the combined h_4 scale to be centrally uni-modal.

Correctness of matching as a similarity rating response, ψ_5 : For each triad order the $t-s-r$ rating ($1 \leq \text{similarity} \leq 5$) was recorded, with half the ratings corrected for laterality so that '1' = 'same as the left-hand one' is now replaced by '1' = 'perfectly correct match'. (The left-hand stimulus is the same as the middle one in half the triads and the same as the right-hand one in the other half.) The $s.d.$ of the pooled frequency distribution of these corrected ratings was used to standardize the 12 mean ratings, for each triad order and stimulus mixture condition, about the mid-point of the similarity rating scale; so that a positive value indicates correct matching. As for h_4 each standardized value (distributed as $(0, 1)$) has a constant added to it for clarity of meaning.

IV. RESULTS

We set out the results in tables, for each of the response measures in turn; the obtained values are given first and the predicted corresponding value in brackets immediately afterwards.

The predicted values are obtained from the relation between the theoretical values in Table I. The linearity assumptions we have made together with the assumption $\psi(0)=0$, or $\psi(0) \ll \psi(\min: \pi, \delta)$ permit us to add the four values of $h_1(W)$ in the second column of Table I headed 'identical pair' [or $h_2(W)$ analogously] to give a sum equal to $h_1(2\delta A + \delta A + 2\delta A + \delta A) = h_1(6\delta A) = 6h_1(\delta A)$, so $\frac{1}{6}\sum h_1(W)$ is our estimate $\hat{\delta}$ of $h_1(\delta A)$, and the predicted values are then proportional to $2\hat{\delta}, \hat{\delta}, 2\hat{\delta}$ and $\hat{\delta}$ reading downwards.

In the case of h_3 and h_4 the predicted values are obtained from the relations between the theoretical values in the third column of Table I, again using our linearity assumptions to combine the four terms, their average being an estimate of $h_3(\pi A)$ [or $h_4(\pi A)$ analogously], for $h_3(\pi - \delta + \pi + 2\delta + \pi + \delta + \pi - 2\delta) = h_3(4\pi)$, so $\frac{1}{4}\sum h_3(W) = \hat{\pi}$ whilst $h_3[(\pi + 2\delta)A] + h_3[(\pi + \delta)A] - h_3[(\pi - \delta)A] - h_3[(\pi - 2\delta)A] = h_3(6\delta A) = 6\hat{\delta}$ as for h_1 or h_2 . We substitute $\hat{\pi}$ and $\hat{\delta}$ values to give the expected values shown in brackets. In the case of ψ_5 we solve as for h_3 or h_4 but with ψ instead of h .

For any measure we thus obtain theoretical values proportional to the corresponding terms in Table I, and with the same mean value for a block as the observed value. The number of observations per cell in the Tables II-VI is 36 for the 9/1 and 7/3 mixtures, and 54 for the 8/2 mixtures but the method of

fitting theoretical values is based on pooled data over all subjects and leaves one degree of freedom for each column of four entries.

The statistical significance of the agreement between observed and theoretical values is not readily assessed as the five measures are not homogeneous in their derivation or psychological scaling type. To facilitate comparison we have calculated the variance of each column of four obtained figures, and the mean squared difference of observed and theoretical values, and expressed these as a ratio of the latter divided by the former: the ratio is obviously zero for a perfect fit and increases with progressively worse theoretical predictions. The variance 'scales up' the mean squared difference for comparisons between measures.¹

Tables II to VI are all arranged on the same plan, as follows: each of the three columns refers to a stimulus mixture condition, first the 90 per cent $A + 10$ per cent B compared with 90 per cent $B + 10$ per cent A , designated 9/1 as in the previous discussion, then the 8/2 mixtures (more similar and comparable than the 9/1) and finally the most similar 7/3 mixtures. We expect π to decrease from left to right in the tables but δ we expect to be constant unless the model is in error and δ is not independent of π , or because of some non-linear distortions in the scaling transformation procedures by which the response measures were obtained δ is not linear on δ . The rows of the tables refer to the four possible orders in which the triads were tasted, using the notation given previously. In each cell the obtained response measure is given, with its theoretical value in brackets after it. At the foot of each column is given the estimated value of δ , and where applicable that of π also, from which estimate (or estimates) the theoretical values immediately above were obtained, the mean squared difference of observed and theoretical values (*m.s.d.*) and the ratio of *m.s.d.* to the variance of observed values (*var. ratio*).

TABLE II. IDENTICAL PAIR PROPORTIONATE PREFERENCES, (h_1)

Normalized observed and theoretical values (in brackets) for order preferences on identical paired comparisons: all positive values are preferences for the second stimulus in the pair tasted:

Order	Mixtures:		
	9/1	8/2	7/3
I	0.18*(0.28)	0.13 (0.13)	0.21 (0.25)
II	0.28 (0.14)	-0.21 (0.07)	-0.08 (0.13)
III	0.18 (0.28)	0.08 (0.13)	0.44 (0.25)
IV	0.18 (0.14)	0.39 (0.07)	0.18 (0.13)
$\hat{h}_1(\delta)$	0.137	0.065	0.125
<i>m.s.d.</i>	0.0103	0.0183	0.0211
<i>var. ratio</i>	5.42	0.403	0.679

The above observed values should be regarded with reservation because of the large proportion of tied preferences [scored $\frac{1}{2}$, $\frac{1}{2}$ in the generally accepted convention] in the raw data.

S.P. ¹cf. Kruskal, J. B. (1964), *Psychometrika*, 29, 9, for a further comment on this.

TABLE III. IDENTICAL PAIR POOLED RATINGS PREFERENCE (h_2)

Standardized observed and theoretical values (in brackets) for order preference as scaled on identical paired comparisons, positive values indicate a mean preference strength greater than average: for the whole data pooled.

Order	Mixtures:		
	9/1	8/2	7/3
I	-0.15 (0.24)	-0.13 (-0.16)	-0.11 (0.04)
II	+0.34 (0.12)	-0.19 (-0.08)	+0.11 (0.02)
III	+0.63 (0.24)	+0.15 (-0.16)	+0.13 (0.04)
IV	-0.11 (0.12)	-0.35 (-0.08)	-0.02 (0.02)
$\hat{h}_2(\delta)$	0.12	-0.08	0.02
<i>m.s.d.</i>	0.0654	0.0455	0.0101
<i>var. ratio</i>	0.622	1.39	1.04

The h_2 values are also to be taken with reservation as the original raw data showed marked response preferences for some statements on the card from which subjects chose a response statement.

TABLE IV. DIFFERENT PAIR PROPORTIONATE PREFERENCE, (h_3)

Normalized observed and theoretical preferences for A or B in different paired comparisons, all positive values are preferences for B .

Order	Mixtures:		
	9/1	8/2	7/3
I	0.67 (0.57)	0.24 (0.32)	-0.03 (0.06)
II	0.81 (0.84)	0.65 (0.68)	0.28 (0.54)
III	0.79 (0.75)	0.59 (0.56)	0.63 (0.38)
IV	0.36 (0.48)	0.24 (0.20)	0.00 (-0.10)
$\hat{h}_3(\pi)$	0.66	0.43	0.22
$\hat{h}_3(\delta)$	0.09	0.12	0.16
<i>m.s.d.</i>	0.0067	0.0025	0.0353
<i>var. ratio</i>	0.207	0.67	0.499

TABLE V. DIFFERENT PAIR POOLED RATINGS PREFERENCE (h_4)

Standardized observed and theoretical preferences for A or B as scaled on different paired comparisons, all positive values are preferences for B , the scale being standardized about the mid-point.

Order	Mixtures:		
	9/1	8/2	7/3
I	0.67 (0.59)	0.22 (0.26)	0.09 (0.11)
II	0.65 (0.74)	0.45 (0.50)	0.41 (0.47)
III	0.79 (0.69)	0.47 (0.42)	0.41 (0.35)
IV	0.44 (0.54)	0.23 (0.18)	0.00 (-0.01)
$\hat{h}_4(\pi)$	0.64	0.34	0.23
$\hat{h}_4(\delta)$	0.05	0.08	0.12
<i>m.s.d.</i>	0.0086	0.0023	0.0019
<i>var. ratio</i>	0.541	0.163	0.055

TABLE VI. CORRECTNESS OF MATCHING AS A SIMILARITY RATING RESPONSE (ψ_5)
 Standardized observed and theoretical t - s - r matching scores, all positive values indicate mean matchings are on the correct side of the scale mid-point.

Order	Mixtures:		
	9/1	8/2	7/3
I	0.71 (0.67)	0.54 (0.61)	0.11 (0.09)
II	0.92 (0.82)	0.81 (0.82)	0.50 (0.48)
III	0.67 (0.77)	0.76 (0.75)	0.33 (0.35)
IV	0.50 (0.62)	0.61 (0.54)	-0.06 (-0.04)
$\psi_5(\pi)$	0.72	0.68	0.22
$\psi_5(\delta)$	0.05	0.07	0.13
m.s.d.	0.0092	0.0025	0.0004
var. ratio	0.411	0.210	0.009

The correlations between corresponding values in the above tables for h_3 and ψ_5 have been calculated; for the observed values $r = +.80$, and for the theoretical values $r = +.98$. One would expect, over the range of differences encountered here, that $h(\psi) = a\psi + b$ where a and b are constants, though some loss of agreement would arise because subjects are not unanimous in their preferences; in other words h may be positively or negatively monotonic on ψ and in any sample of tasters we are faced with the problem of expressing a population of relations of the above form in one equation. The dominant preference for B mixtures here enables us to do this, but there is necessarily some loss of information incurred if we do it.

V. DISCUSSION

The above results in Tables II-VI indicate that the linear trace theory accounts reasonably well for differences between the four presentation orders used. The best correspondence of theory with experimental findings occurs on the responses to stimulus comparisons between physically different stimuli (Tables IV, V and VI), but not apparently between the most different stimuli, the 9/1 and 1/9 Barsac-Graves mixtures. Two aspects of the results deserve comment: the almost complete failure to fit the differences between presentation orders for the identical paired comparisons (Tables II and III), and the very close fit to the 8/2 and 7/3 mixtures columns in Tables IV, V and VI.

The human judge is in no position, on available introspective evidence, phenomenologically to decompose a sensation difference into trace decay (δ) and real difference (π) components; the perceived differences which yield the h_1 and h_2 preferences of Tables II and III are to him merely very small differences, nearer to zero than most, but not all, of the other differences encountered in the test situation. For such small differences we suggest that responses are generated by a decision process which is comparatively inefficient in the sense that it yields too many false positive or negative responses: in the language of signal detection theory which may be more pertinent here than classical psycho-physical threshold models [9] we are working at stimulus differences which are

so slightly above the background variable level of irrelevant taste sensations that the subject's likelihood ratio criterion for reporting any difference is only met part of the time. Our linear trace decay model is in fact a more elaborate form of a classical difference detection theory and cannot by its nature cope with the form of guessing strategy used by subjects responding to what are traditionally considered to be near-liminal or subliminal stimuli. Anomalous results are often reported in this region [25] and we hope soon to publish new data on this problem.

The slightly worse fit of the theory to the more extreme differences in the 9/1 columns of Tables IV, V and VI we consider arises in the following manner. It was noted previously that the two wines differed in other qualities than sweetness. The more extreme mixtures will therefore tend to show differences in other qualities than sweetness because weak taste or odour characteristics of one wine could be just detectable in 9 parts in 10 but just not detectable in 1 part in 10, whereas changing to 8/2 and 2/8 mixtures could make the same minor characteristic apparent in both mixtures. In this way, the 9/1 comparisons need not be made on the same basis by all subjects (neither for preference nor similarity judgements) whilst the 8/2 and 7/2 mixtures would tend to be judged on sweetness alone, as predominant characteristic, and hence comparisons would be on a basis more nearly consistent with the theory. The possibility of multi-criterion discriminations for taste is well known to workers in this field [36] and can be investigated by appropriate multidimensional scaling methods.

Previous work [14] suggests that a linear trace decay model should be appropriate for describing gustatory judgements and affective judgements [4]. Our findings are thus compatible with these earlier studies, but our paradigm is more complex and drawn from an applied rather than a fundamental context.

The poor fit of the model for h_1 and h_2 measures raises the questions (i) how could the model be modified to accommodate the deviant observations, short of replacing it with a signal detection theory which rests on different assumptions, and (ii) what implications has the pertinence of the model when we come to select statistical tests to investigate the non-randomness or psycho-physical causality of taste discrimination and preference behaviour; does it cast doubt on the legitimacy of any currently advocated statistical procedures?

The following suggestions are offered as heuristic. (i) The deviant observations of h_1 and h_2 must be a consequence of trace decay and other response processes superimposed, such as guessing strategies, response sets to the rating scales, and so on. As we are comparing four such identical paired comparisons it would seem reasonable to assume, if we do not wish to discount the between presentation orders variability as random error, that the differences are due to differences in the rate of trace decay, all other sources of response variability being comparable in their effect over the four presentation orders. Lauenstein [24] suggested that the rate of trace decay is a function of the difference between the stimulus intensities encountered and the background level of sensory

excitation, a suggestion which anticipates later adaptation-level theory. Revising our assumptions of the theory to incorporate Lauenstein's idea, we make $\delta B = \lambda \delta A$ where λ is some multiplier such that $1 \leq \lambda \leq U$ (an arbitrary upper bound). There are two triads to consider: for AAB (or BAA), our new decay factor = $\delta' = [(2 + \lambda)\delta]/3$, whilst for ABB (or BBA), $\delta' = [(1 + 2\lambda)\delta]/3$, assuming in each case that δ' is based on an arithmetic mean stimulus against zero background. Writing δ' for δ in the terms of Table I, column one, we obtain Table VII.

TABLE VII. COMPARISON OF TRACE DECAY TERMS FOR IDENTICAL PAIRED COMPARISONS IN ORIGINAL AND MODIFIED THEORY

Tasting order	Original model from Table I	Modified model
I	$h(2\delta A)$	$h[\frac{2}{3}\delta(2+\lambda)A]$
II	$h(\delta A)$	$h[\frac{1}{3}\delta(1+2\lambda)A]$
III	$h(2\delta A)$	$h[\frac{2}{3}\delta(1+2\lambda)A]$
IV	$h(\delta A)$	$h[\frac{1}{3}\delta(2+\lambda)A]$

It will be seen by examination that the main values causing a bad fit of the original model for h_1 and h_2 are those pertaining to order III, where the obtained values are much larger than the theoretical. This discrepancy might be accommodated by making λ large, but it may equally well be the case that a non-linear trace decay model, or one in which serial interference between judgements is allowed for, could accommodate the data. In view of the imprecision in the original response measures it is otiose to consider any more complex theories. (ii) The evidence we have offered casts doubts on the appropriateness of statistical procedures used to analyse the results of duo-trio or triangle tests (with the important possible exceptions of [29, 34]) because we have a judgement process which is not determined by one response parameter, a probability of detecting a difference, which for a given procedure is greater or even less than the chance probability of discrimination, say $\frac{1}{2}$ or $\frac{1}{3}$, but rather a two parameter (π, δ) or a three parameter (π, δ, λ) situation, in which δ need not be negligible relative to π , nor apparently is δ independent of π . This is not in itself a great practical problem because one would hope to operate with differences where π swamped δ , but the point arises that preferences do not map onto π alone, but onto π and δ , and if $h'[\psi(\pi, \delta)]$ is not linear on $\psi(\pi, \delta)$ then some quite marked, reliable and yet artefactual preferences could be reported [16].

VI. GENERALIZATION

It is frequently recommended that subjects chosen for food assessment tasting panels should be screened on their absolute taste thresholds for the four primary tastes; sweet, acid, bitter and salt, before these subjects are employed in tasting experiments [6]. This screening is intended to eliminate insensitive or atypical subjects. This does not meet all the implications of our trace decay

model, for we have introduced a second parameter δ , and whilst it is plausible to assume that π may be a function of individual difference thresholds, we should make no such *a priori* assumptions about δ .

As an indication of what is implied in sampling a population of tasters in the light of the π, δ model, we first assume that $\log \pi$ and $\log \delta$ are distributed as a bivariate distribution [1], then $\log(\pi/\delta)$ will be distributed normally with parameters $\mu_{\pi\delta}$ and $\sigma_{\pi\delta}$.

For a population of judges for whom h and ψ are in $1 \leftrightarrow 1$ correspondence, a representative sample from that population for tasting purposes is one for which $E(\log \pi)$, $E(\log \delta)$ and $E[\log(\pi/\delta)]^2$ are matched respectively with μ_π , μ_δ , and $(\mu_{\pi\delta}^2 + \sigma_{\pi\delta}^2)$ where μ_π , μ_δ are parameters of the parent bivariate population.

We suggest it is in this above sense that a panel of tasters should be shown to be representative for a given gustatory prediction.

VII. CONCLUSIONS

Detailed examination of the judgements in a taste assessment procedure casts doubts on the legitimacy of some statistical models currently used. A linear trace decay model of comparison judgements provides a more valid description and explanation of the results of the assessment procedure, but necessarily complicates the picture. We conclude that a psychophysical examination of taste assessments is to be preferred on grounds of validity to the arbitrary and static descriptions underlying 'biometric' approaches, and that our statistics should attempt to accommodate the psychophysical evidence before any generality can be claimed for quoted results on the discriminability or relative acceptability of tastes.

REFERENCES

- [1] ARRCHISON, J. and BROWN, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press.
- [2] BAKER, G. A., AMERINE, M. A., and ROESSLER, E. B. (1954). Errors of the second kind in organoleptic difference testing. *Food Res.*, **19**, 206-210.
- [3] BAKER, G. A., AMERINE, M. A., ROESSLER, E. B., and FILIPELLO, F. (1960). The nonspecificity of differences in taste testing for preference. *Food Res.*, **25**, 810-816.
- [4] BEEBE-CENTER, J. G. (1932). *The Psychology of Pleasantness and Unpleasantness*. New York: Van Nostrand.
- [5] BENSON, P. H. (1955). A model for the analysis of consumer preference and an exploratory test. *J. appl. Psychol.*, **39**, 375-381.
- [6] BOGG, M. M. and HANSON, H. L. (1949). Analysis of foods by sensory difference tests. *Advances in Food Res.*, **2**, 220-258.
- [7] BORAK, J. (1922). Über die Empfindlichkeit für Gewichtsunterschiede bei abnehmender Reizstärke. *Psychol. Forsch.*, **1**, 374-387.
- [8] BRADLEY, R. A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics*, **9**, 22-38.
- [9] BROADBENT, D. E. and GREGORY, MARGARET (1963). Vigilance considered as a statistical decision. *Brit. J. Psychol.*, **54**, 309-323.
- [10] BYER, A. J., and ABRAMS, D. (1953). A comparison of the triangular and two-sample taste methods. *Food Technology*, **7**, 185-187.

- [11] FERRIS, G. E. (1957). A modified latin square design for taste-testing. *Food Res.*, **22**, 251-258.
- [12] FERRIS, G. E. (1958). The *K*-visit method of consumer testing. *Biometrics*, **14**, 39-49.
- [13] FERRIS, G. E. (1960). A new model for consumer testing. *Food Res.*, **25**, 802.
- [14] FODOR, K., and HAPPICH, L. (1922). Der Bedeutung der Zeit Zwischen Zwei Vergleichreizen bei Bestimmung von Unterschiedsschwelle. *Arch. ges. Physiol.*, **197**.
- [15] GOLDIAMOND, I. (1958). Indicators of Perception. *Psychol. Bull.*, **55**, 373-411.
- [16] GREGSON, R. A. M. (1960). Bias in the measurement of food preferences by triangular tests. *Occup. Psychol.*, **34**, 249-257.
- [17] GREGSON, R. A. M. (1963). Validation problems in interpreting preference responses to mixed food qualities. *Applied Statistics*, **12**, 1-13.
- [18] GRIDGEMAN, N. T. (1959). Sensory item sorting. *Biometrics*, **15**, 298-306.
- [19] GRIDGEMAN, N. T. (1964). Sensory comparisons: the 2-stage triangle test with sample variability. *J. Food Sci.*, **29**, 112-117.
- [20] HARPER, R. (1952). *Psychological and Psychophysical aspects of Craftsmanship in Dairying*. *Brit. J. Psychol. Monog. Supp.*, **28**, Cambridge University Press.
- [21] HELM, E., and TROLLE, B. (1946). *Wallerstein Lab. Commun.*, **9**, No. 28, 181.
- [22] KAMENETSKY, J. (1959). Contrast and convergence effects in the ratings of foods. *J. appl. Psychol.*, **43**, 47-52.
- [23] KÖHLER, W. (1923). Zur Theorie des Sukzessivvergleichs und der Zeitfehler. *Psych. Forsch.*, **4**, 115-175.
- [24] LAUENSTEIN, O. (1933). Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler. *Psych. Forsch.*, **17**, 130-177.
- [25] LINKER, E., MOORE, MARY E., and GALANTER, E. (1964). Taste thresholds, detection models, and disparate results. *J. exp. Psychol.*, **67**, 59-66.
- [26] NEEDHAM, J. G. (1935). The effect of the time interval upon the time error at different intensive levels. *J. exp. Psychol.*, **18**, 530-543.
- [27] PERYAM, D. R. (1958). Sensory difference tests. (Chapter 3 in Little, A. D., *Flavour Research and Food Acceptance*, New York: Reinhold Pub. Corp.).
- [28] SCHUTZ, H. E., and BRADLEY, J. E. (1954). Effect of bias on preference in the difference-preference test, in Peryam, D. R., Pilgrim, E. J., and Peterson, M. S., *Food Acceptance Testing Methodology*, Advisory Board on Quartermaster Research and Development Committee on Foods, U.S. National Academy of Sciences.
- [29] SHEFFÉ, H. (1952). An analysis of variance for paired comparisons. *J. Amer. statist. Ass.*, **47**, 381-400.
- [30] TANNER, W. P., Jr., and SWETS, J. A. (1954). The human use of information. *IRE Trans. on Information Theory*. Vol. PGIT-4.
- [31] THURSTONE, L. L. (1950). Methods of food-tasting experiments in *Proceedings of the second conference on research*. American Meat Institute, 85-182.
- [32] TITCHENER, E. B. (1912). Description versus statement of meaning. *Amer. J. Psychol.*, **23**, 165-182.
- [33] TORGESSON, W. S. (1958). *Theory and Methods of Scaling*. New York: John Wiley and Sons.
- [34] URA, S. (1957). On Scheffé's analysis of variance for paired comparisons. *Rep. statist. Appl. Res.*, JUSE, **4**, 132-146.
- [35] URA, S. (1960). Pair, triangle and duo-trio test. *Rep. statist. Appl. Res.*, JUSE, **7**, 107-119.
- [36] YOSHIDA, M. (1963). Similarity among different kinds of taste near the threshold concentration. *Jap. J. Psychol.*, **34**, 25-35. (In Japanese, with English summary).



HIGHER-ORDER FACTORS

By C. J. ADCOCK

University of Illinois

The paper discusses the hierarchical conception of factorial structure developed by Burt and his research students, and examines the alternative approaches preferred by Thurstone and Cattell. It is argued that second-order factors cannot be distinguished from first-order factors by mathematical criteria. It follows that interpretational criteria are requisite and the method of transformation described by Schmid and Leiman is recommended.

I. INTRODUCTION

The hierarchical conception of factorial structure has two sources. The first and earliest derives from the centroid method of analysis—the method of ‘simple summation’, as it was originally called—and is associated with the name of Cyril Burt (1940, 1950). He has argued that a hierarchical concept of human mind is plausible on neurological grounds, and that centroid factors provide a confirmatory scheme in so far as the first factor is usually positive throughout and all succeeding factors are bipolar, thus permitting repeated subdivision and capable of alternative representation by ‘basic’ and ‘group’ factors (cf. Moursy, 1952). To such a scheme there are familiar objections. In this case all the factors fall within an arbitrary reference frame determined by their initial representation in the sample of tests; the number of levels found depends upon the number of factors; and the scheme must in practice be compressed into a highly artificial model irrespective of actual causal factors, unless (as Burt insists) the tests are carefully chosen with some hypothesis in view. Otherwise the results from any given particular sample of variables can only be used to gain some preliminary insight into possible structure and can never provide a fixed scheme which can be replicated. Burt (1954) has taken note of these criticisms, and meets them by maintaining that the function of the statistical analysis is primarily to confirm hypothetical causal structures tentatively inferred on the basis of antecedent evidence.

The second line of approach begins with a non-hierarchical or reticular concept, derived by oblique rotation. Thurstone suggested criteria for getting away from the arbitrary reference frame of centroid analysis, and demonstrated that such criteria could be meaningful provided the factors were oblique. The principle of rotation to oblique simple structure has been widely accepted. It entails, however, the recognition of second and even higher-order factors if we are to account for all of the variance, and so in the end we get back to a hierarchical model, but one which enables us to determine the number of levels by explicit criteria and the position of our reference frame at each level by simple structure and/or psychological criteria.

Cattell (1962) has argued against both hierarchical concepts. He regards higher-order factors as evidence of causes acting on the lower-order factors, but insists that this does not imply a hierarchy, although the factor model may seem to indicate it. The seeming hierarchy is only part of a wider system which takes the form of a network or reticule. Empirical findings often seem to accord with such a view. The present writer is inclined to think that, although the causal relations will always be complex, and often manifest themselves in a correlational network, yet we shall be able to distinguish hierarchical systems whose interconnections will explain the reticule. Such a hope can do no more than generate hypotheses, however. In the empirical case we can expect to find a reticule. Our interpretation of these relations may or may not suggest important hierarchies at work; but we cannot beg the question by assuming either that they must be there or that they cannot be there.

In this paper I wish to put forward a strong plea for flexibility in interpreting factorial data. It is only too easy to set up a mathematical model, base a technique upon such a model, and then, because we have obtained results in this way, assume that the data must be in accord with the model assumed. Obviously to allow the more or less arbitrary choice of method to determine the range of our hypotheses is to prejudge every situation. I shall illustrate this by reference to the concept of higher-order factors itself.

II. HIGHER-ORDER FACTORS

When is a Factor a Higher-Order Factor? The definition of a higher-order factor here adopted is simple enough. It is a factor whose variance extends over the range of several lower-order factors, but always in proportion to the weighted sum of the lower-order loadings.¹ Normally we indicate this variance only in terms of the loadings of the higher order on the lower order; but since this variance is measured through the variables obtained in the first place it is also possible to report it in terms of the variables themselves as described, for example, by Schmid and Leiman (1957).

In Table I, I give data for a very simple case involving only two primaries and one second-order equivalent. In this case the second-order factor has equal loadings on both primaries, and these are non-overlapping; so that all the

TABLE I. A GENERAL FACTOR WHICH IS EQUIVALENT TO A SECOND-ORDER FACTOR.

A	B	C
40	80	
30	60	
20	40	
30		60
25		50
20		40

¹ In this *Journal* the definition of a 'higher-order factor' is that of Thurstone and Cattell. The type of factor here described is what Spearman and Holzinger originally termed a 'bifactor'. Later this latter term was used as described below, and the proportional factors were termed 'bifid', see this *Journal*, 3, p. 59 and refs. (*Editorial Note*).

relations are obvious on inspection. No one of these second-order equivalent loadings may be changed without also making a corresponding change in the primary involved, else it will cease to conform to the second-order model. On the other hand, it is quite feasible to have a general factor which does not have this relation to the group factors. Table II illustrates a parallel form in which the proportionality required by the second-order factor is not met.

TABLE II. A GENERAL FACTOR WHICH IS NOT EQUIVALENT TO A SECOND-ORDER FACTOR

A	B	C
60	80	
50	60	
30	40	
50		60
40		50
30		40

A factor pattern of this kind can be obtained either by Burt's (1950) group-factor approach or by Holzinger's bifactor method (Harman, 1960). Neither method has had widespread use, because they lack the flexibility of the oblique rotated centroid of principal axis solutions. Both aim at non-overlapping group factors, and both allow only one general factor. Only in specially selected cases is this likely to fit the data. In most cases we can expect overlapping group factors, wider factors spanning several group factors and even more than one general factor. I am not, therefore, advocating these methods as against the oblique rotation approach. All we are concerned with here is the relation between a general factor and the group factors which it spans. The difficulty of calculating such a group factor, except in very simple instances, is irrelevant. The factor pattern is theoretically possible; and, provided the data are suitable, can be calculated. What then does it mean?

Clearly such a general factor differs from a second-order equivalent in that its variance is not mediated by the group factors or the primaries. Spearman proposed an interesting analogy for his general factor, although he himself was unaware of the distinction we are considering. His notion of a common source of energy operating through various motors involves just the mathematical relation required by the second-order model. Changes in power-supply can affect work-output only through the motors concerned, i.e., through the primaries. In contrast to this we might consider the case of income which is determined partly by a man's age and partly by the salary range obtaining in his profession. If the former is assessed and received directly without relation to his employment, we have the effect of a general factor not geared to the group factors represented by the several professions. In theory it might be argued that Spearman should not on his own principles have entertained the power-motor analogy with regard to his general factor, nor can we validly adopt the age-profession analogy for a second-order factor. Neither conception fits the model; and all interpretation along lines such as these should, strictly speaking, be excluded.

In practice, however, it would be exceedingly foolish to do so. The same data may give rise to either kind of factor. Admittedly the data of Table II cannot be squeezed into the form of Table I; but, provided we are prepared to tolerate a little fuzziness (which in the empirical case will be present in *both* solutions) these are genuine alternatives depending solely on the method of analysis. The second-order case will generally involve at least one additional factor; but none of these factors will necessarily prove fully significant. Harman's (1960) 24-variable example is resolved into a general factor and five group factors by the bifactor method. Only four of the group factors turn up in his oblique rotated centroid solution; but inspection of the latent roots of the principal factor solution shows that by Guttman's criterion he should have taken out a further factor; and, had he done so, the rotated outcome would have been similar to the result obtained by the centroid factors based on the original bifactor solution, which does in fact yield corresponding group factors. This example is an excellent illustration of the fact that the same data will give parallel results when handled by different methods; and the decision as to which the model is to be assumed (other than for convenience of calculation) is one that cannot be based on mathematical criteria as such.

Since mathematical criteria do not appear to offer us any safe means of distinguishing between data which need to be interpreted in terms of a second-order model and data which call for a general factor model, we are thrown back on other criteria. The psychological meaning will be crucial. Any finding must be fitted into a wider psychological system, and all we can do is to accept the interpretation which best fits the wider framework. This means that we must always, irrespective of the computational method we have chosen, consider both possibilities. But it means also that we can usually choose our method on the basis of its computational advantages, and may therefore select the more flexible oblique rotation approach. By submitting such a solution to a Schmid-Leiman (1957) or Cattell-White (1963) transformation we may gain the advantage of seeing the higher-order variance as it relates directly to the variables as well as in the usual form. This enables us to give adequate consideration to the general-factor hypothesis; but it also has the merit, in the Schmid-Leiman case, of providing us with coefficients which will indicate the relative importance of the factor in regard to the variables concerned. As Cattell has emphasized, the usual oblique factor pattern does not provide true loadings in terms of which the correlations might be generated. With the Schmid-Leiman (1957) transformation we have all variance represented in a single level orthogonal system, the number of factors being the sum of those appearing at all levels and the orthogonality of the factors deriving from the fact that all correlation between factors is now replaced by the variance of the higher-order equivalents.

The drawback to such a procedure is that the lower-order factors at each stage are credited only with the unique variance they contribute directly, and their contribution via the higher order factors appears separately. Cattell and White (1963) wish to have their cake and eat it too; and consequently they

adopt a compromise which retains the full variance of the higher-orders¹ at the expense of no longer having an orthogonal system permitting inter-level comparisons. For some purposes this will certainly prove an advantage; and there is room for both methods of approach. But it is necessary first of all to make sure that one's choice of method does not prejudge the issue of interpretation.

In the light of this discussion it seems possible to adopt some simple rules of procedure. In view of the restrictions of the bifactor method it would seem desirable in most cases to begin with a principal factor solution, using estimated communalities in the diagonal and possibly iterating several times to improve the communalities, the number of factors being first decided by the Guttman criterion or by some estimate of sampling error. At this stage we can carry out an oblique rotation, and proceed to higher-order analysis. Finally we carry out a Schmid-Leiman transformation. The last form becomes the basis of our interpretation, but should not bias our attitude. We must take account of the hypothesis suggested by the higher-order relation. There must be a constant reference to the original higher-order analysis and possibly also to a Cattell-White transformation. Such a procedure permits us to bring out all the significant relations without committing ourselves to any particular model for our interpretations.

III. INTERPRETATION OF GENERAL AND HIGHER-ORDER FACTORS

I shall now consider certain hypotheses which may be advanced to account for what appears as a higher-order factor in oblique analysis and which *may* derive from some kind of general factor. Let us begin with what is probably the most frequent case, that of a true higher-order factor as required by the mathematical model.

A True Higher-Order Factor. The simplest case will be that of a single second-order factor. The level is of no significance since we can always measure a higher-order factor directly by making use of appropriate tests.² It may appear at different levels in different analyses. What we are concerned with here is the nature of the relation between the different levels.

We may take as an instructive example the box analysis of Thurstone (1947) and Thomson (1961). Here the three primaries represent length, breadth, and depth. In this case there is only one hypothesis we can advance for the second-order factor, namely, that it arises from the restriction placed on the range of dimensions in the boxes actually used. Since boxes are always produced for some special purpose and are made of some special material which in itself limits possibilities, the more extreme variants (e.g. half inch high by one mile long) will never arise. We thus get a necessary correlation between length, breadth, and depth due to sample restriction. This must always be mediated by the

¹ What Cattell calls the "associated factor variance" as opposed to the "disassociated factor variance."

² One merit of the Cattell-White transformation is that it parallels the variance which would be obtained by direct measurement at the higher-order level.

primary factors, so that we have here a true second-order relation. And notice that we are forced to this conclusion even if we have obtained our results by the bifactor or group factor method. Obviously a general factor can be interpreted by second-order analogy.

It will be useful here to consider some implications arising out of the transformation proposed by Schmid and Leiman. This apportions the variance between the contributions arising from the group factors of length, breadth, and depth, and the contribution arising from the sampling restriction. This is here a useful distinction. It is one thing to deal with variables because they involve the common dimensions of length, breadth, and depth, and another to deal with them as related to one another by reason of the concomitance of these three dimensions. For certain purposes these covariances are what primarily concern us, and the sampling effect is just noise to be eliminated. Whether we should concern ourselves with the 'unique' variances (Schmid and Leiman, 1957) or with the 'total' variances for each factor must depend on circumstances. For studying theoretical implications the former will usually be needed; for practical applications the latter may be preferable.

To what extent does this second-order interpretation apply to Spearman's ' g '? We might suggest that some at least of the variance might derive from a concomitance of aptitudes genetically derived and tending to affect sometimes more than special aptitudes or faculties. Despite the old adage, beauty and brains are positively correlated; and it is quite likely that some of the ' g ' effect is akin to the general 'boxiness.' Spearman's 'cognitive energy' would also be an example of second-order explanation; but there is little or no evidence that it is more than an analogy. Rate of neurological functioning could produce similar effects if we always use speeded tests; but refinement in testing should separate anything like this as a distinct speed factor. Complexity of the neurological system as a whole (Burt's theory) might also be considered, and could be compared to the parallel increase in capacity among electronic computers where again several functions will work better in the newer and high-capacity computers; but this again is really similar to the 'boxiness' concept. The parts tend to improve concomitantly.

A True General Factor. With a true general factor we have a source of variance which spans the variables involved in several group factors, but without any necessary relation to their variance. The general factor is itself an independent source of variance. We have already suggested an illustration. A similar situation might exist if a group of people whom we were examining for foreign language efficiency had developed this by two main methods—study in their respective schools and varying lengths of time spent in the foreign country concerned. Assuming that the time spent abroad was not a function of the school they had attended, it would probably emerge as a general factor, and the schools would emerge as group factors.

Possibly some part of the Spearman's ' g ' could be explained in this way. The notion of intelligence as insight might also suggest a general factor of this

kind. No matter what task he may be concerned with the intelligent person may find some way to boost his output. A capacity to see relations in any area sufficiently familiar could give rise to a factor of a general rather than a second-order type, and could best be conceived in terms of its unique variance.

Vicarious Functioning. In human functioning we have to take note of an important phenomenon, namely, vicarious functioning. This occurs in a variety of forms. Consider the cube test which has been used to measure capacity for visualization. If one has vivid, well controlled visual imagery, one can mentally enact the sectioning of a painted cube into 27 smaller cubes, and make the requisite observations with regard to the number of painted sides; but persons who fail to do the task by visual imagery may achieve excellent results by mathematical reasoning. Again a person who has difficulty in the verbal application of the laws of deductive logic may compensate effectively by making use of Euler's circles, and so substitute visual for verbal capacity. To the extent that this sort of thing occurs we may expect a general factor to be generated. This again may make some contribution to the Spearman's 'g'. It is allied to the insight effect. In fact, we might almost define insight as the factor which has the maximum capacity for vicarious functioning and for instigating other forms of vicarious functioning.

Overlapping Effects. The hypothesis I now wish to present arises from a consideration of personality analysis, but I shall illustrate it with a purely hypothetical case, so that the reasonableness of the hypothesis can be assessed independently of whether the empirical evidence supports it or not. Let us assume that we have evidence for four traits: aggressiveness, shyness, affectionateness, and group dependency. We refrain from defining shyness as the antithesis of aggressiveness, since we may have a complete absence of shyness without aggression and complete absence of aggression without shyness. Affectionateness and group dependency will both be manifested in a desire for social contact and will so have similar effects resulting in behavioural correlation. Similarly shyness and aggressiveness will be correlated through their social effects. Either aggressiveness or lack of shyness will reduce inhibition in social relations, but not in the same way. Lack of shyness may facilitate the satisfaction of the social needs manifested in affectionateness and group dependency; aggressiveness has antithetical effects. The correlation pattern which thus results can be represented as in Table III.

TABLE III. CORRELATION PATTERN ILLUSTRATING THE OVERLAPPING OF EFFECTS
Significant positive coefficients are indicated by *X*;
non-significant coefficients by *O*.

Variables	1	2	3	4
1. Aggressiveness	—	<i>X</i>	<i>O</i>	<i>O</i>
2. Freedom from Shyness	<i>X</i>	—	<i>X</i>	<i>X</i>
3. Affectionateness	<i>O</i>	<i>X</i>	—	<i>O</i>
4. Group dependency	<i>O</i>	<i>X</i>	<i>O</i>	—

The question we have to ask is what happens when such a pattern of primary factor correlations turns up. There can be no doubt that variables two to four would be involved in a general factor; and it is quite likely that aggressiveness would pick up some loading in the average matrix where it will probably have some small correlation with both three and four rather than the zero indicated here. Yet in what sense can we postulate that there is any common causation operating on even variables two to four? It looks much more like a common effect than a common cause. Affectionateness could be conceived as predisposing towards group dependency, but need shyness have any direct relationship to these two? Yet if we use measures involving social interaction, we shall probably get correlation in terms of the common result; and, since all the traits necessarily involve social interaction of some kind, we can hardly avoid introducing such effects. Here then is another hypothesis which must be kept in mind when we are interpreting general and second-order factors. Whether the particular case we have presented is empirically confirmed is a matter for actual investigation and does not concern us here.

IV. SUMMARY AND CONCLUSIONS

I have tried to demonstrate that a second-order factor cannot be distinguished from a basic or general first-order factor by mathematical criteria since the same data may yield either type of solution according to the method chosen. We are therefore thrown back on interpretational criteria. This being so, it can be argued that an oblique rotation of principal factors transformed, as advocated by Schmid and Leiman (1957), provides the most convenient representation. I have further referred to four types of interpretation to illustrate the range of the hypotheses which need to be considered in dealing with higher-order factors. In particular I have sought to show that a higher-order factor does not necessarily have to be equated with a 'common cause.'

REFERENCES

- BURT, C. (1940). *The Factors of the Mind: An Introduction to Factor Analysis in Psychology*. New York: MacMillan.
- BURT, C. (1950). Alternative methods of factor analysis. *Brit. J. statist. Psychol.*, 2, 98-121.
- BURT, C. (1954). The sign pattern of factor matrices. *Brit. J. statist. Psychol.*, 7, 15-29.
- CATTELL, R. B. (1962). The basis of recognition and interpretation of factors. *Educ. psychol. Measmt.*, 22, 667-697.
- CATTELL, R. B. and WHITE, O. (1963). The use of higher-order personality factors in relation to variables. In Press.
- HARMAN, H. H. (1960). *Modern Factor Analysis*. Chicago: University of Chicago Press.
- MOURSY, E. M. (1952). The hierarchical organization of cognitive levels. *Brit. J. statist. Psychol.*, 5, 151-180.
- SCHMID, J. and LEIMAN, J. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22, 53-61.
- THOMSON, G. H. (1961). *The Factorial Analysis of Human Ability*, 5th Ed. New York: Houghton Mifflin Co.
- THURSTONE, L. L. (1947). *Multiple-Factor Analysis*. Chicago: University of Chicago Press.

THEORETICAL CONCEPTS IN PSYCHOLOGY

By RICHARD W. COAN

University of Arizona

A more comprehensive scheme for classifying theoretical concepts is needed to clarify the issues involved in choosing concepts for psychology. The scheme here proposed concentrates on distinctions of observational and ontological status which seem basic to such issues. Psychology can best advance as a science through a more liberal acceptance of various types of concepts whose use is at present too often rejected, notably concepts embodying existential hypotheses, or designating so-called 'private' phenomena of consciousness.

I. THE CLASSIFICATION OF CONCEPTS

In recent years psychologists have shown considerable concern with the legitimacy of various concepts employed in their theories. Much of the discussion has centred round a distinction made by MacCorquodale and Meehl [27] between 'intervening variables' and 'hypothetical constructs'. Its importance cannot be questioned; but a good deal of confusion has arisen owing to a failure to recognize the limitations of this dichotomy. The issues usually discussed involve a number of conceptual distinctions, some of which were not incorporated by MacCorquodale and Meehl in their classification. It is the contention of the present writer that many of the important issues relating to psychological concepts can best be treated in the context of a more comprehensive scheme; and it is the purpose of this paper to present such a scheme and then examine the problems so raised.

MacCorquodale and Meehl distinguished between (i) variables definable and quantifiable in terms of specified manipulations of empirical variables and (ii) concepts not reducible to empirical terms. Their recommendations succeeded to some extent in standardizing current usage. A number of comparable distinctions have been made by other writers, e.g. Carnap's [10] 'disposition terms' and 'theoretical terms'. Objections have been raised with respect to the proposed dichotomy, in particular by Bergmann [2, 3]. A distinction made on the basis of reducibility versus non-reducibility to observables meets with certain difficulties. According to the assumptions and definitions one chooses to adopt, it can be argued either that nothing is really observed directly and that all the entities and processes of which we speak are inferences from our experience, or that all the meaning of our concepts and statements is ultimately reducible to immediate observation, or that all concepts contain a certain amount of

irreducible surplus content in addition to a core reducible to observation. On any of these grounds it would follow that no firm distinction between hypothetical constructs and intervening variables can be maintained. Yet most psychologists seem content to assume that it has some practical value, even if it is to be regarded simply in terms of relative position along a continuum of abstraction or remoteness from empirical operations.

Restatements or modifications of the MacCorquodale-Meehl distinction have been proposed by such writers as Ginsberg [18] and Rozeboom [36]. The literature also contains suggestions for subcategorization within the two classes, as well as proposals that cut across the distinction. Rozeboom [36] points out that hypothetical constructs may vary with respect to type of underlying speculation and with respect to the level of uncertainty regarding their content. Alternative schemes of classification within the construct realm are O'Neil's [32] classification of hypotheses and classifications of models presented by such writers as Bertalanffy [4] and Zubin [41]. Other types of "models"—as the term is defined by Braithwaite [6]—are included in a somewhat more comprehensive scheme used by Argyle [1] in enumerating different types of theories in social psychology.

It is obviously possible to classify theoretical concepts along many different lines. Indeed, in a less formal way psychologists often make typal distinctions not incorporated into any of the schemes cited above, and certainly concepts can be classified with respect to either ontological status, factual content, mode of experience, source of experience, or the manner in which the concept is introduced into theory. A man's epistemological and metaphysical assumptions naturally affect his preferences for methods of classification: phenomenists and realists will prefer different schemes, and the scheme preferred by a logician will be declared useless by a psychologist. It is, in short, futile to argue that any one mode of classification is intrinsically more valid than another. We must therefore be content to classify psychological concepts along lines that reflect the issues which face psychology at the present stage in its history.

In the writer's opinion a classificatory scheme that best serves the current needs of psychology must be constructed in terms of observational and ontological status. Accordingly, the scheme proposed will now be presented in outline as a guide to the discussion that follows. The specific distinctions made within it will be clarified later on when discussing the various issues relevant to the classification. The most fundamental distinctions may not be immediately apparent, since the writer has used a novel terminology to avoid words that have acquired conflicting usages. Within the scheme the main categories are capable of further subdivision and cross-classification should anyone wish to add further distinctions. In proposing this scheme of theoretical concepts, the writer has conceived of a *theoretical concept* in a very broad sense. Thus the scheme does not rest on any prior distinction between theory and model, between theory and law, or between theoretical terms and observation terms. It is intended broadly

to encompass the domain of concepts used in scientific description and explanation in psychology. The proposed scheme is as follows:

- I. Adspective concepts
 - A. Omnispective concepts
 - 1. Classificatory omnispective concepts
 - 2. Relational omnispective concepts
 - B. Propriospective concepts
 - 1. Classificatory propriospective concepts
 - 2. Relational propriospective concepts
- II. Ultraspective concepts
 - A. Hypothetical concepts
 - 1. Specified hypothetical concepts
 - 2. Unspecified hypothetical concepts
 - B. Fictional concepts

II. THE OBSERVATIONAL BASIS OF CONCEPTS

The fundamental distinction used in this scheme pertains to the observational foundation of the concept. An adspective concept is one definable in terms of observed entities, events, or relations. An ultraspective concept would not be so definable, although it might consist of components that are. In short, the ultraspective concept contains "surplus meaning" in the sense of factual content that goes beyond past observation. Thus, the concept *unicorn* is ultraspective. The constituents of the concept are traceable to past observation, but their composite is not. Likewise an electronic brain model would consist of ultraspective concepts, even though parallel concepts in the field of electronics might be regarded as adspective. In general, the "surplus content" might involve any of the three levels of uncertainty for the existence of a hypothetical construct discussed by Rozeboom [36], or it might take a suppositional form in the case of a model adopted for convenience. It involves something which has not been observed with respect to the objects to which it is applied, but which may or may not be assumed to be ultimately observable.

We have already noted some of the difficulties in using observational status as a basis for classification, but in the opinion of the writer this mode of classification is necessary. In any empirical science, one seeks to make sense of accumulated observations and to move in the direction of fresh observations, and the theorist must make discriminations along the dimension represented here by adspective and ultraspective concepts.

It is not assumed that a sharp dividing line can be drawn between adspective and ultraspective concepts. In a sense every concept contains surplus meaning, for even the simplest communicable concept embodies considerable abstraction and induction which carry it far beyond any immediate observation. From a phenomenological standpoint, one can argue that every concept is actually a construct in so far as it represents an inference regarding a physical reality underlying our experience. In any case, the ultraspective

concept is viewed here as one whose surplus content, from any philosophical viewpoint, exceeds the bounds of mere abstraction or generalization from past observations. Outside an intermediate range, the distinction between adspective and ultraspective concepts should not be difficult to apply.

It should be noted that this and other distinctions are made here on the basis of the denotative content accorded a concept by the theorist in his explicit definition and use of it. Any adspective concept may have considerable surplus connotative content. This connotative content, though irrelevant to present concerns, may arouse needless controversy when misconstrued as denotative content and may subtly affect the course of scientific pursuits as it gradually becomes denotative. Kendler [23] has noted the tendency for controversies in the field of learning theory to arise because of a failure to distinguish between operational meanings and the "intuitive properties" ascribed to the concepts utilized by such men as Tolman and Hull. Meissner [31] and others have pointed to a rather pervasive tendency toward a reification of theoretical concepts and introduction of surplus meaning that has previously been formally excluded. Much of this process undoubtedly involves a gradual conversion of connotative meaning to denotative meaning, and while it is conducive to certain difficulties in communication, it may also be fundamental to scientific progress.

Most of the controversy regarding the relative virtues of intervening variables and hypothetical constructs has centred about the adspective-ultraspective dimension. There are those like Marx [29] who regard operationally valid intervening variables as "the only kinds of constructs ultimately admissible in sound scientific theory". They may feel, as he does, that hypothetical constructs are tolerable as a temporary expedient, or they may advocate a strict exclusion of speculative content at all times. Others, like Maze [30], see a danger in the use of the intervening variable. Maze feels that it is likely to be perceived as providing explanation or extending knowledge when it actually provides only a restatement of fact. Still another viewpoint is that of Hebb [20], who favours the use of hypothetical constructs, but believes that mere "tautological constructs" may serve a valuable purpose.

The general trend in recent psychology has apparently been toward greater acceptance of hypothetical constructs and a recognition of the vital role that they play. While many still feel that we should strive toward theories that will ultimately be devoid of ultraspective concepts, there is a growing suspicion that the notion of an ultimate science or an ultimate theory may be illusory. Furthermore, there seems little doubt that whatever the merits of reconceptualizing ultraspective concepts as adspective concepts, science progresses as much by the opposite process. Bertalanffy's [4] examples of the transformation of formal into material models are at least illustrative of this; and Feigl [14] has presented arguments in favour of converting intervening variables into hypothetical constructs. Perhaps scientific progress may best be regarded as involving a helical trend in which surplus factual content is repeatedly infused and later expelled or accepted

as fact on the basis of accumulating observations. It might be argued that a science that ceases to move in this manner by constantly reaching beyond the limits of extant evidence has ceased to fulfil one of its basic functions.

We may note that the recent emphasis on the intervening variable and the later and more liberal advocacy of hypothetical constructs has been paralleled by shifts in attitudes toward operationism. The operationist position, of course, is subject to many shades of interpretation. In its more liberal forms, it is quite consistent with the use of hypothetical constructs as long as they point fairly clearly in the direction of confirming operations that have not yet been performed. In recent years, however, there has been a growing tendency to permit the scientist to employ any kind of concept he sees fit to use. Logicians like Carnap [10] have pointed out the dangers of regarding empirical meaning as totally reducible to operational definition. While the value of operationism is generally recognized, there has been increasing recognition of the inherent absurdity and potentially stifling effects of a strict operationism—both noted by Ginsberg [19].

III. VARIETIES OF ULTRASPECTIVE CONCEPTS

Closely linked to the issue of adspective versus ultraspective concepts are a variety of issues relating to the choice of ultraspective concepts. The basic distinction drawn here is between *hypothetical* and *fictional* ultraspective concepts. In the hypothetical concept the surplus factual content refers to entities, events, or relations that are regarded as potentially observable, whether or not the means of observation are available. The fictional concept, on the other hand, refers to entities, events, or relations which are assumed not to correspond to fact and to be unobservable either currently or potentially. The fictional concept commonly assumes an analogical form, in that the things to which reference is made have been observed in a different realm and provide an understanding by way of a comparison. Many so-called 'models' are composed of fictional concepts. The hypothetical *construct* is sometimes construed as the equivalent of what is here called a hypothetical *concept*; but by the definition of MacCorquodale and Meehl [27] it encompasses the fictional concept as well.

A further distinction within the class of hypothetical concepts seems desirable; and it is proposed that they be subdivided into specified and unspecified hypothetical concepts. A specified hypothetical concept is one for which the definition contains a relatively specific statement of the properties that must be observed to provide confirmation. In the case of the unspecified concept there would be minimal indication of these properties; and the hypothetical element would be confined to a general indication of the type of properties to be sought or of the region in which they might be found. This distinction is essentially the same as that of O'Neil's [32] 'characterized' and 'uncharacterized' hypothetical terms. Thus, a 'memory trace' defined as some unspecified alteration of nervous tissue, would constitute an unspecified hypothetical concept. The unspecified hypothetical concept bears some kinship to the formal physiological concepts advocated by Pratt [33], to Lindzey's [26] conventional construct,

and to the constituents of Bertalanffy's [4] formal models; but it is not strictly equivalent to any of these.

Specified and unspecified concepts represent regions along a continuum of explicitness, rather than completely separate and discrete categories. At the unspecified end of the continuum, hypothetical concepts shade into adspective concepts in which the hypothetical element is strictly connotative. Since the intentions of psychological theorists are often obscure, many borderline cases will be encountered. A similar problem arises with fictional concepts, since it is often difficult to determine whether a theorist intends explicitly to adopt a model concept, or merely to preserve a model concept in his choice of words and exclude it by formal definition from denotative reference.

In discussions of the relative merits of hypothetical and fictional concepts, it is the fictional concept whose value is most often questioned. MacCorquodale and Meehl [27] assert that hypothetical constructs "should not assume inner events that cannot possibly be true." They thus recognize the value of what is here called the hypothetical concept, but they reject the use of fictional concepts. In particular they object to the incorporation of certain psychoanalytic concepts, such as *libido* and *censor*, into physical models. Krech [25] presents a similar viewpoint in objecting to notions of existing structures that are non-behavioural, non-experiential, and non-neural. Bridgman [7], Zubin [41], and Braithwaite [6] have spoken of the virtues of models in a way which shows they see a potential heuristic value in the use of fictional concepts. At the same time, they recognize certain dangers inherent in their use. Bridgman and Braithwaite stress the difficulties arising from literal acceptance of the model as a physical reality, while Zubin points out the stagnation resulting from the use of a model that is not sufficiently testable to allow its own possible demise.

In his early pleas for the use of existential hypotheses, Feigl [12, 13] advocated the use of hypothetical concepts in general, though he pointed out the risks of using those requiring operations remote from present technology. On the other hand, he explicitly rejected the use of concepts unconfirmable in principle, citing animistic and vitalistic concepts as examples. He has also commented (Feigl [14]), that certain fictional concepts may have great heuristic value, as with Freudian analogical models. What appears important in his standpoint is that the fiction be recognized as such, so that we are prepared to replace it with existential hypotheses when the time is ripe. The basic defect in the concepts criticized by Feigl—the later ether hypothesis, soul, entelechies, etc.—is that they are improperly accorded an assumed or hypothetical existential status. If we rely on the intention of the conceptualizer, they must be classed as unconfirmable hypothetical concepts, or, in some cases, adspective concepts.

A different position is presented by Mandler and Kessen [28], who argue that the existential status of a model concept is irrelevant to its value within theory. *Libido*, whatever hydraulic properties we assign to it, meets the requirements of sound theory if it yields testable and confirmed statements regarding

the behaviour it is intended to explain. Presumably Mandler and Kessen would criticize some of the concepts that Feigl criticizes, but chiefly on the ground of their inability to generate testable statements, rather than their unconfirmable existential status.

Another issue that has attracted attention is the kind of factual content that should be embodied in hypothetical concepts. Hypothetical concepts of physiological content have been advocated by such writers as Pratt [33, 34], Tolman [39], Krech [25], Klein and Krech [24], and Hebb [20]. A few have equated the hypothetical construct with physiological hypotheses. The formulation of MacCorquodale and Meehl [27], however, implies no such restriction; and it would be contrary to the present classification to impose it on the use of the term *hypothetical concept*. Furthermore, I should agree with such writers as Lindzey [26] and Rozeboom [36] that, whatever value physiological hypotheses may have, they do not constitute a course that all psychologists *must* follow. In psychology the surplus content of useful hypothetical concepts may relate to behavioural events, experiential events, social processes, or various kinds of functional relations. Unless we make *explanation* and *reduction* synonymous by definition, it is absurd to assume that the physiological route is the only one to take in seeking a better explanation. If 'explanation' is to be defined in terms of stated functional relations in general, it would be presumptuous to suppose that reductive explanation—i.e., explanation in terms of correlations between physiological and psychological events—is the only kind still lacking. To take but one area in which we might fruitfully deal with behavioural hypothetical concepts, there is the problem of inborn behavioural predispositions: here our knowledge of the bare facts is still at a primitive level; and, despite the ostrich-like manoeuvres of some behaviourists, the problem still persists.

Germane to the choice of physiological or non-physiological concepts is the question of ultimate physiological reduction as a goal of psychology. With Lindzey [26], Rozeboom [36], and Mandler and Kessen [28], the writer would agree that there is no basis for assuming that physiological explanation will inevitably be the most effective kind (in terms of its capacity to generate predictions), nor will it necessarily serve the same purposes as non-physiological explanation. As Hebb [21] has recently noted, physiological concepts cannot take the place of psychological ones. Along with the increased interest in physiologizing, there is a growing feeling that the goal to be sought is not a mere reduction of psychological to physiological formulations, but a greater mutual translatability of psychological and physiological formulations, and the construction of theories comprehending both physiological and behavioural events.

IV. VARIETIES OF ADSPECTIVE CONCEPTS

The classification of adspective concepts presents special problems, since it depends on judgements regarding the nature of observation. As Hempel [22] has noted, the concept of *observation* may be construed in a variety of ways. It may be construed broadly, or limited to what is publicly ascertainable. Because

of the problems inherent in the concept of observation, some of the issues considered here to lie within the realm of adspective concepts are often confused with issues involving the adspective-ultraspective dimension.

The object sought in classification of concepts was to introduce distinctions which, on the one hand, would be conducive to unambiguous and comprehensive categorization of concepts, and at the same time would be relevant to the issues on which psychologists are forced to make decisions. With respect to varieties of observation, the basic issue is one commonly represented by such elastic terms as public versus private, objective versus subjective, behavioural versus experiential, etc. Unfortunately, despite a core of meaning that seems to run through these terms, a host of overlapping distinctions is bound up with them as they are commonly applied. The need for introducing novel terms if we are to focus on a well defined and useful distinction is obvious.

It would be possible to formulate a distinction between omnispектив and propiospective concepts in a variety of ways, and there are special problems imbedded in all of them. Obvious distinctions between the internal and the external or peripheral prove to be irrelevant, whether the skin or the ego-boundary is taken as a dividing line. The usual formulations lead us to think primarily of the social observability or confirmability of the entity embodied in the concept. The latter is not helpful; for if it is taken to imply repeatability by more than one observer, it must be present in every psychological concept. More fundamental than mere duplication of verbal report is the question whether the thing observed and reported is the same for any two observers. We are inclined to assume that it is the same when we speak of a "chair" or a "prehensile movement", and that it is not the same when we speak of a "body image" or "feeling of grief". The fact with which we are forced to reckon, however, is that there is no operational means by which we can distinguish directly between the two presumed realms of experience. Though we can compare subsequent behaviour, verbal or otherwise, we can never compare the actual experiences or observations of two people so as to establish their identity. Nevertheless, it seems important to distinguish between the table that we both see and the pain that we experience as individuals, even though our experiences in both instances are inevitably private. The important point is that, despite commonsense presuppositions, the distinction cannot be made in terms of the conditions of observation; it must rather be made on the basis of conceptual or theoretical operations. Thus, as the distinction is formulated here, the omnispектив concept is one whose observational content is conceptualized as public; and the propiospective concept as one whose observational content is regarded as subject to individual observation only.

This distinction rests on a realistic basis so far as it assumes distinction between experience and the object experienced. To a strict phenomenalist, it may appear rather pointless, although it can also be expressed in terms of conceptual operations applied to phenomenal objects. In any case, as long as

the bulk of psychological theory continues to rely on essentially realistic assumptions, the categories here proposed may have useful applications. Within the framework of present-day psychology, it is possible to apply them either according to the evident intentions of the formulator of a concept or according to the consensus of a number of persons who use it. The writer, however, has no wish to defend the thesis that this distinction corresponds to any sharp dividing line within "reality" itself.

Having distinguished two types of adspective concepts, the present scheme further distinguishes under each of them the two subtypes classificatory and relational. The classificatory concepts are those for which the defining operations are confined to abstraction and classification of things and events; they include proper names and their equivalents as single-member classes. Thus they embrace much of the same realm that various philosophers speak of in terms of observation language. As commonly used, *chair*, *neurone*, *John Smith*, *Methodist*, *arm movement*, and *red object* are classificatory omnispective concepts; *pain*, *anger*, *phenomenal self*, and *green after-image* are classificatory proprioceptive concepts. In both groups we have concepts representing various degrees of abstraction.

The relational concept is one in which the defining operations exceed abstraction and classification; they introduce a relation or conjoint function involving two or more of the things or events designated by classificatory terms. It is thus definable in terms of classificatory concepts, and may ultimately be related to extant observation in the same way as the classificatory concept. The relational function may assume any of a variety of forms: e.g., an interactive form (referring to a product of a specified quantitative combination of things or events) or a conditional form (as in the case of a term designating a property whose expression in an object is dependent on the state of other objects or events). The common examples of intervening variables or disposition concepts would thus be relational omnispective concepts. Comparable possibilities exist for the proprioceptive realm. For example, concepts defined in terms of predispositions whose manifestation would be experiential, such as *guilt proneness*, may be regarded as instances of relational proprioceptive concepts. Many psychological concepts, however, are defined in terms of a mixture of omnispective and proprioceptive components, and must thus be considered either mixed classificatory or mixed relational concepts.

If we are interested in logical completeness of classification, then it should be noted that the omnispective-proprioceptive dichotomy cuts across the adspective-ultraspective dimension, and is potentially applicable to the ultraspective realm as well. It would seem possible to subclassify hypothetical concepts as omnispective and proprioceptive, according to the assumed nature of the unobserved event. This is not proposed here, because the practical value of the resulting subcategories seems questionable, and would lead to problems that do not arise when the omnispective-proprioceptive distinction is applied

to adspective concepts only. It seems likely that a slightly different distinction—such as that which Smith [37] makes between objective and subjective constructs—would prove more useful in the sphere of hypothetical concepts.

As the hypothetical concept is here defined, its content could obviously refer to entities or events conceptualized as observable either publicly or only privately. Examples of the former—unobserved planets, unobserved neural structures—come readily to mind. Within the body of personality theory there are concepts that might be construed as examples of the latter sort. In Jungian theory, for instance, we find such things as 'archetypal' concepts and the 'self', which can be defined in terms of experiences that most people have not had, but could have. In general, however, a theorist does not formulate such concepts unless the potential observations lie within the range of past observations. The concepts themselves might then be regarded as adspective.

Unfortunately, we commonly experience greater difficulty in communications about things we consider privately observable than we do about things we consider publicly observable. Our rudimentary concepts seem less precise; and we assume a more critical attitude toward reports of things we have not observed ourselves if the things described lie in the private realm. Thus, we may not hesitate to accept the report of a single astronomer regarding a previously unreported celestial body; yet to a British or American psychologist the concurring reports of many continental phenomenologists seem to contain a great deal of hypothetical, if not meaningless, content. Since the difference between the public and the private is not to be found in the conditions of observation *per se*, consistent classification seems to demand that we should classify a concept as adspective whenever it is defined in terms of observations that someone has made.

It is probably correct to say that behaviourism, more than any other movement in psychology, has brought attention to bear on the omniscientive-propriospектив distinction, and raised the question of the values of these two classes of adspective concepts for science. The term *behaviourism*, however, embraces a great variety of viewpoints; and, as Feigl [16] has noted, it is capable of at least three different interpretations—materialism, logical behaviourism, and methodological behaviourism. Materialism denies the existence of "raw feels", and would presumably lead to a rejection of propriospектив concepts as devoid of meaning. Logical behaviourism insists on defining concepts in terms of physical observation, and it implicitly denies the omniscientive-propriospектив distinction, since it maintains that a so-called propriospектив term is meaningful only so far as it is reducible to omniscientive terms. Methodological behaviourism leaves open the question of the existence of "raw feels", but it holds that only those types of behaviour which are expressible in omniscientive terms are scientific subject matter. Combinations of these three positions are obviously conceivable. Psychologists who support the behaviourist movement generally favour a *methodological behaviourism*; philosophers who support it more often stress *logical behaviourism*, e.g. Bergmann [2] and Carnap [11]. Carnap, however,

describes his position as physicalism rather than as behaviourism, and is clearly critical of methodological behaviourism (Carnap [10]). Feigl [16], on the other hand, considers all three unacceptable.

A common extension of logical behaviourism is that which advocates the use of public corollaries of private events either as operational equivalents or as behavioural indicators of experience: e.g. Boring's proposal [5] to reduce experience to discrimination, and recent discussions about the use of verbal reports. Such an approach may help to broaden the scope of behaviouristic psychology; but its dangers have been stressed by Pratt [35] and Zener [40].

Many writers, including Tolman [38], Pratt [33, 35] and Feigl [12], have pointed out that all sciences are ultimately rooted in private experience. Hence, psychology must likewise rest on the private experiences of psychologists and their experimental subjects. If we accept this view and yet regard science as a public enterprise, we must identify its public character with something other than its ultimate source of data. This is done in various ways; but almost inevitably it is viewed in terms of the communicative aspects. Some writers, like Carnap [11], focus attention on science as a logical system of sentences. Others are more concerned with social concordance. Pratt [33, 35] stresses reproducibility of report as the hallmark of science; and for him reproducibility requires a statement of operations that will permit other investigators to duplicate the observations and so confirm or refute our own report. He excludes from the category of scientific subject-matter "experiences which can be had by one person only", but not because they are private: private experiences such as those studied by early introspectionists and rejected by behaviourists are in his view scientifically respectable, just because they do lend themselves to reproducible reports.

Others, like Braithwaite [6] and Zener [40], also hold that the public character of the data is not an essential requisite. Braithwaite instead emphasizes the application of hypothetico-deductive method. Zener stresses the "repeatability of obtained functional relationships between specified experiences and specified conditions (external and internal)". A similar idea seems implicit in Feigl's suggestion [17] that all sorts of mental states and events will be found to fit neatly into the framework of experimental science and are able to satisfy our demands for confirmable statements, provided we view them as components in a general network of lawfully related facts.

At present there appears a growing consensus to the effect that, whatever demands science may make of the theorist, the circumscription of subject matter imposed by behaviourism is not one of them. Even throughout the era in which behaviourism seems to have cornered the market for textbook definitions of psychology, psychology has never really ceased to be at least partly a science of experience. The recent pleas by Burt [8], Zener [40], Feigl [17], and Hebb [21] form the more conspicuous crests in a tide that has long been rising. From a strict behaviourist standpoint it would seem that the scientist should

confine his attention to the afferent impulses that arise from his exteroceptors, preferably those reaching the brain by way of the second cranial nerve. The fact remains, however, that man possesses rich and widely varied experiences, for many of which there seems no adequate vehicle of interpersonal communication. If the psychologist is not sensitive to these experiences, he can hardly hope to deal with them scientifically at all. The writer fully agrees with Zener [40] that to advance in our scientific treatment psychology will need more "experientially sensitive observers". Hitherto we have successfully kept many out and desensitized such observers among our graduate students by rigorous methodological training.

It could be argued that, since any definition of science is arbitrary, science ought to be construed in such a way as to exclude a direct concern with experience. If so, then it should be granted that science furnishes a poor foundation for any application concerned with problems of human existence. If the psychotherapist is to view his client as something more than an object that needs a bit of conditioning or synaptic reassortment, he will need training in a philosophical co-discipline that does concern itself with the appropriate subject matter.

V. EVALUATIVE CRITERIA IN SCIENCE

Many of the issues that have been discussed revolve around questions regarding the criteria that a theorist should employ in making differentially evaluative judgments. Some can be clarified if we focus on evaluative criteria themselves. The criteria that have received greatest attention are those relating to the acceptance of theoretical formulations. Though the distinctions are somewhat arbitrary, it is convenient to think of them as falling into three groups: syntactic, semantic, and pragmatic. The syntactic criteria are concerned with the structure of the theory as a symbolic system; the semantic criteria with the relation between this system and the realm of observations as data to which it relates; and the pragmatic criteria with the effect of the theory on scientists—i.e., with the theory's heuristic value. Thus all commonly advocated criteria are encompassed in the following list:

I. Syntactic criteria: A. Simplicity, conceptual economy; B. Internal consistency.

II. Semantic criteria: A. Testability, verifiability, or confirmability; B. Explicitness, operational clarity of concepts; C. Incorporation of known findings; D. Comprehensiveness.

III. Pragmatic criteria: A. Immediate research productiveness; B. Theoretical productiveness.

Such a list could be reorganized in different ways, for the criteria are variously interrelated; and there is little agreement regarding the weight that should be attached to any of the criteria. Indeed many theorists regard some of them as actually irrelevant. On this abstract level we tend to regard the pragmatic criteria as not germane to the problem of judging the goodness of a theory.

Yet our judgements of theories are probably tempered by a knowledge of their influence. If we base our judgements solely on syntactic and semantic criteria, there is still room for extensive variation; and if we regard a theory not as a mirror held up to nature, but as a structure superimposed on nature for human purposes, it becomes obvious that the personal characteristics of the theorist must largely determine such things as the relative emphasis on descriptive and predictive precision or simplicity.

The mood of British and American theorists would lead us to emphasize chiefly semantic criteria, especially the first two listed above. Unfortunately it is with respect to these two that there is the greatest uncertainty. The first, which we have called testability, verifiability, or confirmability, relates to the truth value of a theory. It is now generally agreed that truth as such is not a criterion generally applicable to scientific theories. Psychologists commonly insist that a theory should be verifiable—i.e., that it should yield predictions whose conformity to fact can be directly determined. On the other hand, philosophers like Carnap [9, 10] and Feigl [15, 17] have argued that theories are best viewed as conceptual systems possessed of free construction and only indirectly tied to underlying observations. They thus propose the somewhat weaker criterion of confirmability. This is more widely applicable, for it permits a graduation of judgements with respect to theories manifesting little direct verifiability. With this criterion we might perceive merits in certain ultraspective formulations that would otherwise be rejected. The concept of testability refers simply to the possibility of performing relevant empirical operations, and may be viewed as complementary to the other two.

The criterion listed above as IIB concerns the demands to be placed on the ingredients of the scientific vocabulary. It could be formulated in terms of the clarity of the relation between the theoretical term and observation; and, where a choice is possible, we should no doubt prefer a concept manifesting this kind of clarity to a high degree. It is possible, however, that the desired characteristic is better assessed in terms of stability of usage than by analysis of definitions. Mandler and Kessen [28] insist that the basic requisite of scientific concepts is “reliable usage . . . in particular observational circumstances and in the sentences of a theory”.

Invariance of usage is a syntactic characteristic akin to internal consistency; but it is a likely consequence of what is here called explicitness. And certainly it may furnish a more broadly applicable standard than explicitness. It may be difficult to evaluate the operational clarity of many propriospective terms by an analysis of definitions, but reliability of usage is indirectly evidenced in the concordance of formulations containing these terms.

Some of the issues discussed in preceding sections concern a choice among the criteria listed above. In the case of the question of adspective versus ultraspective concepts, several semantic and pragmatic criteria are involved. Ultraspective concepts can probably be best defended on the grounds that they

afford comprehensiveness and heuristic value. An avoidance of such concepts is most likely to be maintained on grounds allied to verifiability and explicitness. Adspective concepts will usually fare better in the light of these two criteria. But there is still great variation in explicitness within both the adspective and the ultraspective realms; and an unverified concept may be highly verifiable.

The more specific question of the value of fictional concepts involves the same issues. The usual defence for such concepts is their heuristic value; and the objections commonly raised against ultraspective concepts are most often levelled at fictional concepts. However, it is clear that one's final judgement regarding the merits of a theory depends not only on the weight attached to the criteria, but also on the particular way the criteria are applied. As Mandler and Kessen treat the first two semantic criteria, a fictional concept might be judged quite favourably. A fictional concept is inherently unverifiable; but they argue that the criterion should be applied to theoretical statements containing the concept rather than to the existential status of the concept itself. Verifiability, they imply, is not separable from reliability of usage, which may be high for a fictional concept, even though operational clarity must be low.

Turning to omnispective versus propriospective concepts, we find that the problems of evaluation become more complex. Several different kinds of evaluation are involved; and these have seldom been properly separated. We need to distinguish at least three questions. The first is the one on which we have focused attention—the acceptance of theoretical formulations; a second is the acceptance of subject matter for scientific investigation; a third, the acceptance of observational data as usable or useful.

The second is the most fundamental, since it relates to the nature and purpose of science itself. We could argue that no criteria should be devised for the acceptance of subject matter. The history of science suggests that its purpose is to bring systematic order into all the contents of human experience. There is no area of experience which scientists have universally agreed to exclude. Traditional taboos or theological doctrines may have delayed the entry of certain areas. Development has usually begun earliest, and proceeded most rapidly, in areas least concerned with man's own place in the universe. But the history of science is an account of an ever-expanding subject matter. Psychology is on an outer frontier, and psychologists have sometimes drawn their own boundary lines, lest others should doubt that they are part of the continent of science.

The criterion regarded as basic for observational data is repeatability. To be of use to the scientist, this criterion must be flexible. If we accept no restrictions on subject matter, we can demand repeatability of observations only to the extent that it is possible to obtain it. Physical scientists in fields like astronomy and geology have used data of low repeatability. The carefully controlled experiment insures a high level of repeatability; but it is not a *sine qua non*. An overemphasis on rigid controls can result in a neglect of problems involving complex inter-relations and inadequate observation of events that are

transformed by the very imposition of control. In other fields the choice of method follows the choice of subject matter. In American psychology there is the danger that an initial choice of method will be used as a basis for restricting subject matter. A similar danger arises when certain criteria for evaluating theories are rigidly applied. The criterion of verifiability may be applied in comparing two theories designed to account for the same events. But this too may often become a basis for choosing subject matter, regardless of the fact that various subjects may not lend themselves equally well to verifiable theories. A restriction of subject matter has been widely advocated by behaviourists. This may lead to respectable research; yet it runs counter to the general course of science. It is commonly defended on grounds of repeatability and verifiability. But the extension of such criteria to the choice of subject matter is unjustified. Overt movement, physiological processes, and inner experience constitute three sources of observation, and three realms of subject matter. Each warrants scientific investigation and an attempt at explanation. The last may involve events which are no more controllable or repeatable than earthquakes, and may yield formulations which are no more verifiable than a celestial event observed by a single lonely astronomer. The fact that it challenges the ingenuity of the investigator should not constitute a permanent barrier in the path of intellectual progress.

REFERENCES

- [1] ARGYLE, M. (1957). *The Scientific Study of Social Behaviour*. London: Methuen.
- [2] BERGMANN, G. (1951). The logic of psychological concepts. *Phil. Sci.*, 18, 93-110.
- [3] BERGMANN, G. (1953). Theoretical psychology. *Ann. Rev. Psychol.*, 4, 435-458.
- [4] BERTALANFFY, L. von (1951). Theoretical models in biology and psychology. *J. Pers.*, 20, 24-38.
- [5] BORING, E. G. (1945). The use of operational definitions in science. *Psychol. Rev.*, 52, 243-245.
- [6] BRAITHWAITE, R. B. (1953). *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*. Cambridge: Cambridge Univer. Press.
- [7] BRIDGMAN, P. W. (1928). *The Logic of Modern Physics*. New York: MacMillan.
- [8] BURT, C. (1958). Definitions and scientific method in psychology. *Brit. J. statist. Psychol.*, 11, 31-69.
- [9] CARNAP, R. (1936). Testability and meaning. *Phil. Sci.*, 3, 420-468.
- [10] CARNAP, R. (1956). The methodological character of theoretical concepts. In Feigl, H., and Scriven, M. (Eds.) *Minnesota studies in the philosophy of science*. Vol. I. Minneapolis: Univer. of Minnesota Press. Pp. 38-76.
- [11] CARNAP, R. (1959). Psychology in physical language. In Ayer, A. J. (Ed.) *Logical Positivism*. Glencoe, Illinois: Free Press. Pp. 165-198.
- [12] FEIGL, H. (1945). Operationism and scientific method. *Psychol. Rev.*, 52, 250-259.
- [13] FEIGL, H. (1950). Existential hypotheses. *Phil. Sci.*, 17, 35-62.
- [14] FEIGL, H. (1951). Principles and problems of theory construction in psychology. In Dennis, W., et al. *Current trends in psychological theory*. Pittsburgh: University of Pittsburgh Press. Pp. 179-213.
- [15] FEIGL, H. (1956). Some major issues and developments in the philosophy of science of logical empiricism. In Feigl, H., and Scriven, M. (Eds.) *Minnesota studies in the philosophy of science*. Vol. I. Minneapolis: Univer. of Minnesota Press. Pp. 3-37.

- [16] FEIGL, H. (1958). The 'mental' and the 'physical'. In Feigl, H., Scriven, M., and Maxwell, G. (Eds.) *Minnesota studies in the philosophy of science*. Vol. II. Minneapolis: Univer. of Minnesota Press. Pp. 370-497.
- [17] FEIGL, H. (1959). Philosophical embarrassments of psychology. *Amer. Psychologist*, 14, 115-128.
- [18] GINSBERG, A. (1954). Hypothetical constructs and intervening variables. *Psychol. Rev.*, 61, 119-131.
- [19] GINSBERG, A. (1955). Operational definitions and theories. *J. gen. Psychol.*, 52, 223-245.
- [20] HEBB, D. O. (1951). The role of neurological ideas in psychology. *J. Pers.*, 20, 39-55.
- [21] HEBB, D. O. (1960). The American revolution. *Amer. Psychologist*, 15, 735-745.
- [22] HEMPEL, C. G. (1958). The theoretician's dilemma. In Feigl, H., Scriven, M., and Maxwell, G. (Eds.) *Minnesota studies in the philosophy of science*. Vol. II. Minneapolis: Univer. of Minnesota Press. Pp. 37-98.
- [23] KENDLER, H. H. (1952). "What is learned?"—a theoretical blind alley. *Psychol. Rev.*, 59, 269-277.
- [24] KLEIN, G. S., and KRECH, D. (1951). The problem of personality and its theory. *J. Pers.*, 20, 2-23.
- [25] KRECH, D. (1950). Dynamic systems, psychological fields, and hypothetical constructs. *Psychol. Rev.*, 57, 283-290.
- [26] LINDZEY, G. (1953). Hypothetical constructs, conventional constructs, and the use of physiological data in psychological theory. *Psychiatry*, 16, 27-33.
- [27] MACCORQUODALE, K., and MEEHL, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 55, 95-107.
- [28] MANDLER, G., and KESSEN, W. (1959). *The Language of Psychology*. New York: Wiley.
- [29] MARX, M. H. (1951). Intervening variable or hypothetical construct? *Psychol. Rev.*, 58, 235-247.
- [30] MAZE, J. R. (1954). Do intervening variables intervene? *Psychol. Rev.*, 61, 226-234.
- [31] MEISSNER, W. W. (1960). Intervening constructs—dimensions of controversy. *Psychol. Rev.*, 68, 51-72.
- [32] O'NEIL, W. M. (1953). Hypothetical terms and relations in psychological theorizing. *Brit. J. Psychol.*, 44, 211-220.
- [33] PRATT, C. C. (1939). *The Logic of Modern Psychology*. New York: MacMillan.
- [34] PRATT, C. C. (1941). A reply to Professor Melton. *Psychol. Bull.*, 38, 236-240.
- [35] PRATT, C. C. (1945). Operationism in psychology. *Psychol. Rev.*, 52, 262-269.
- [36] ROZEBOOM, W. W. (1956). Mediation variables in scientific theory. *Psychol. Rev.*, 63, 249-264.
- [37] SMITH, M. B. (1950). The phenomenological approach in personality theory: some critical remarks. *J. abnorm. soc. Psychol.*, 45, 516-522.
- [38] TOLMAN, E. C. (1935). Psychology versus immediate experience. *Phil. Sci.*, 2, 356-380.
- [39] TOLMAN, E. C. (1949). Discussion. *J. Pers.*, 18, 48-50.
- [40] ZENER, K. (1958). The significance of experience of the individual for the science of psychology. In Feigl, H., Scriven, M., and Maxwell, G. (Eds.) *Minnesota studies in the philosophy of science*. Vol. II. Minneapolis: Univer. of Minnesota Press. Pp. 354-369.
- [41] ZUBIN, J. (1952). On the power of models. *J. Pers.*, 20, 430-439.

NOTES AND CORRESPONDENCE

THE STABILITY OF FACTORS

By CYRIL BURT

In one or two early papers I described and illustrated a method of assessing the degree of stability shown by homologous factors obtained in different investigations (Burt, 1939, 1943); and in a later note in this *Journal* (Burt, 1954) I endeavoured to answer certain criticisms advanced by Mr. Leyden (Leyden, 1953). However, both he and other correspondents have now raised a further problem. "Thomson," we are told, "in his discussion of the problem follows much the same Pearsonian procedure as Burt in his 1934 Memorandum, and yet reaches the opposite conclusion. 'An experimenter,' he contends, 'beginning with a *sample* set [of variances and covariances] will not reach the *population* factors'; the factors derived from a particular sample, such as he analyses, will differ widely from those which would be derived from the entire population or from any other sample" (Thomson, 1948, pp. 194 f.).

In point of fact, however, Thomson's procedure is *not* the same as my own. My formulae and my proof, as given in the 'Memorandum on Correlations obtained from Selected Groups' to which Mr. Leyden refers, were generalizations of the method of partial correlation, and were applied to correlation matrices with communalities in the leading diagonal; these were then treated as matrices of variances and covariances. Thomson used Aitken's formulae, and applied them to correlation matrices with units in the leading diagonal. To obtain the factors, I adopted Pearson's method of weighted summation, whereas Thomson commonly uses the centroid method. I assumed several common factors and eliminated specifics; Thomson's analysis assumed only one common factor and included specifics. Finally, my proof did not require the numerical values or even the precise nature of the 'directly selected' variables to be known; Thomson's analysis seems to have given the impression that the values for the selective variables must be known or estimated at the outset.

The results of these differences can be more clearly explained if we consider first of all, not Thomson's imaginary battery of tests, but the actual set of inquiries which gave rise to the whole issue. To facilitate the comparison, however, I shall substitute Thomson's notation for my own. In 1917 I applied a group of scholastic tests to a group of London school children and found four 'common factors'. I assumed that these factors would hold good of the London school population generally and of all the elementary subjects taught in the London schools. From time to time I applied similar tests to other groups; and other S.P.

investigators, Dr. Carey for example, did the same. As a consequence what we all wanted to know was whether the factors so discovered, which *looked* much the same, really *were* the same.

To cope with this problem I imagined a matrix derived by applying a complete set of all conceivable tests applied to a population consisting of all the persons that might conceivably be tested; and I supposed that, in any actual investigation, the battery of tests employed and the group of persons tested were suitably selected samples. The empirical correlation tables thus obtained I regarded as submatrices of the ideally complete matrix of variances and covariances. In our various researches the changes in factor values often appeared too large to be attributable to purely random selection; and the variables which might systematically affect the selection (age, school, social conditions, and doubtless various unknown influences) I called the 'selective variables'. Using Thomson's notation (see his Table, *loc. cit.*, p. 312) we can let R_{pp} represent the correlation matrix for the selective variables; R_{qq} the correlation matrix for a particular sample of tests before selection; and R_{ff} the correlation matrix for the factors. The common factors obtained by factorizing R_{qq} (with communalities in the diagonal) are given by the equation

$$R_{qq} = F_{qf} F'_{qf}, \quad (1)$$

and, since they are principal components, $F'_{qf} F_{qf} = R_{ff}$, the diagonal matrix of latent roots.

I then supposed that, in any given case, the nature of the selection would be indicated by the change of variances and covariances from R_{pp} to V_{pp} (still using Thomson's notation), and that, as result of this selection, R_{qq} would change to V_{qq} , the values actually observed. Now selecting children of a given age can be regarded as a reduction in the variance for age and in the correlations of age with other variables, both selective and non-selective; similarly for the influence of the other selective variables. This reduction can be assessed by generalizing the ordinary formulae for partial correlation. We then find (Pearson, 1912)

$$V_{qq} = R_{qq} - W_{qp}(R_{pq} - V_{pq}) \quad (ii)$$

$$= R_{qq} - W_{qp}(R_{pp} - V_{pp})W'_{qp}, \quad (iii)$$

where the regression coefficient, W_{qp} in my notation and $R_{qp}R_{pp}^{-1}$ in Thomson's formulae (*loc. cit.*, p. 189), reduces to $F_{qf}F_{fp}^{-1}$. (For proof, see Burt, 1943, pp. 9–10, eqns. vii to ix.)

Mutatis mutandis the same equation holds for the modified covariance matrix obtained for the factors, i.e., V_{ff} which takes the place of R_{ff} as a result of selection. Rearranging terms, we then have

$$F_{fp}^{-1}(R_{pp} - V_{pp})(F_{fp}^{-1})' = I - V_{ff}. \quad (iv)$$

And, on substituting this value in eqn. iii, we obtain

$$V_{qq} = F_{qf}V_{ff}F'_{qf}. \quad (v)$$

Evidently, if we factorized V_{qq} , we should obtain as our factor saturations $F_{qf}V_{ff}^{1/2}$. Now if the method of selection happens to ensure that the factors of V_{pp} are proportional to those of R_{pp} , then $V_{ff}^{1/2}$ will be a diagonal matrix, and the factors of V_{qq} will be proportional to those of R_{qq} . Hence the *direction cosines* for the several factors will be identical (my definition of factor invariance). Under similar conditions of selection the same proportionality will hold between the factors of V_{qq} and those obtained in any other sample, e.g., the factors of Thomson's V_{ss} or V_{tt} .

In general, however, V_{ff} will not be diagonal. In that case, as I suggested in this *Journal* (VIII, p. 86), V_{ff} may be factorized by weighted summation: we then obtain $V_{ff}=HDH'$, where H and D denote the latent vectors and latent roots. We can then take

$$\begin{aligned} R_{qq} &= F_{qf} H \cdot H' F'_{qf} \text{ and} \\ V_{qq} &= F_{qf} H D^{1/2} \cdot D^{1/2} H' F'_{qf}. \end{aligned}$$

The rotation of the population factors will thus procure the requisite proportionality; and a similar procedure can be used in the case of the two samples. But to retain proportionality for further samples, we should have to keep on changing the rotations. And these simply furnish us with factors that are essentially dependent on our method of selection. For this reason, I concluded, it is "better to prove or assess similarity than to postulate it and rotate accordingly".

When V_{ff} is a diagonal matrix, then the product of the two covariance matrices $R_{qq}V_{qq}=F_{qf}F'_{qf}$. $F_{qf}V_{ff}F'_{qf}$ will be symmetrical, since, as we have seen, $F'_{qf}F_{qf}$ is also a diagonal matrix. When V_{ff} is not diagonal, then this product matrix will diverge from strict symmetry; and to assess its approximation to symmetry—and therefore the degree of similarity between the two sets of factors—I proposed a 'proportionality criterion', obtained by calculating the unadjusted correlation between the sums of the columns and of the rows respectively (Burt, 1939, pp. 67-69 and Table IV).

Limitations of space have compelled me to give only a brief outline of the approach I have advocated. I have said nothing about the sampling of tests—a still more complex problem; but, if we accept what I called the 'reciprocity principle' and correlate not tests but persons, much the same conclusions will be reached. (For details the reader may refer to the papers cited.) In any case, the chief lesson to be drawn from the foregoing discussion is not the use of any formulae, but rather the need to plan the collection of data in advance, when any factorial analysis is contemplated. Unless there is no better alternative, it is unwise just to take any set of data because they happen to be at hand, factorize them, and then rotate the factors obtained to fit other data, or worse still the investigator's preconceptions.

From the formulae themselves the reader may deduce for himself the principles that should govern the selection of samples. The relevant factor saturations of the selective variables, and the regressions which depend upon them, should be as small as possible; the differences between the covariances

for the selective variables (though not necessarily between these variances) should also be as small as possible; and the two sets of factor saturations involved in these covariances should be as nearly proportional as possible (what this implies in practice I have explained in the papers already quoted). In point of fact these seem to be the principles which most competent investigators have implicitly adopted. Hence, as a study of the literature shows, the factors actually ascertained, at any rate in the field of cognitive abilities, show an unexpected degree of invariance. Even in the imaginary sample worked out by Thomson himself there is a close approximation to proportionality: after selection, the factor saturations derived from the new covariances are 0.33, 0.28, 0.23, 0.19; and the 'unadjusted correlation' is practically perfect; (when derived from the new correlations, they are 0.56, 0.46, 0.37, 0.29, but the 'unadjusted correlation' is still over 0.999).

REFERENCES

- BARLOW, J. A. and BURT, C. (1954). The identification of factors from different experiments. *Brit. J. statist. Psychol.*, 7, 52-56.
- BURT, C. (1939). The relations of educational abilities. *Brit. J. educ. Psychol.*, 9, 45-71.
- BURT, C. (1943). Validating tests for personnel selection. *Brit. J. Psychol.*, 34, 1-19.
- BURT, C. (1955) ap. Cattell, R. B. and Cattell, A. K. S. Factor solutions for proportional profiles. *Brit. J. statist. Psychol.*, 8, 83-92.
- LEYDEN, T. (1953). The identification and invariance of factors. *Brit. J. statist. Psychol.*, 6, 119-120.
- PEARSON, K. (1912). The general theory of the influence of selection on correlation and variation. *Biometrika*, 8, 437-443.
- THOMSON, G. H. (1948). *The Factorial Analysis of Human Ability*. London: University of London Press.

OBITUARY

CHARLES WILFRID VALENTINE

C. W. Valentine grew to manhood in a new era of British psychology. Ward's article in the ninth edition of the *Encyclopaedia Britannica* in 1886 was a watershed. After its publication, psychology moved from the embrace of philosophy and aspired to be a science and mainly an experimental science. In spite of this, in 1900, when Valentine began the study of psychology, there were no chairs of psychology in this country. What is equally important, few looked to psychology for practical help, its applications seemed to most people to be of little account. To its rapid growth and recognition in the last sixty years Valentine made an enduring contribution.

He was born at Runcorn in 1879, and had a varied education proceeding in turn from Nottingham High School to Preston Grammar School, thence to University College, Aberystwyth, where he was an exhibitioner, and to Cambridge, whence he emerged with a double first in philosophy and psychology. He taught for seven years in secondary schools, including a period at St. Olave's, Southwark, and then went to St. Andrews as a lecturer in psychology to the Provincial Committee for the Training of Teachers, and Assistant in Education at the University. It was during this period that he wrote his first book, *An*

Introduction to the Experimental Psychology of Beauty, in the series 'The People's Books', published at Edinburgh in 1913 by T. C. and E. C. Jack. This was the first book in English in its field, and in the preface Valentine thanked Stout for suggesting that he should deal with the psychology of aesthetics.

Promotion was not long delayed; and Valentine accepted the chair of Education in the Queen's University of Belfast. After occupying this for five years, he came to Birmingham in 1919 and stayed in the area for the rest of his long life. Here he taught with distinction: he founded and edited the *British Journal of Educational Psychology*; he wrote widely and influentially; he greatly raised the prestige of education in the University; he was for some time Chairman of the City of Birmingham Higher Education Sub-Committee; and he was head of a department that included a large women's training college.

Valentine did much in Birmingham to establish his department through his reputation as a psychologist and in other ways. He took a keen interest in his advanced students from whose ranks not a few moved to influential positions in educational administration. His lectures to intending teachers were clear and apposite: they were never disquisitions on academic psychology difficult of application; rather they were human in approach and enlivened by *ad hoc* inquiries about fear, curiosity, sympathy and the like. He remembered and was influenced by his own experience as a schoolmaster. In and out of the lecture room he was approachable, interested in his students' personal affairs, and perceptive of ability.

The founding of the *British Journal of Educational Psychology* in 1931 was indicative of the growth in this country of a new field of inquiry. Early this century there were links between education and psychology; and by the thirties these had been greatly strengthened. Valentine's heroic and devoted editorship of the journal over twenty-five years has been described elsewhere. Sir Cyril Burt said in 1956, when Valentine resigned from the editorship, "the journal he started achieved not only financial success but a world-wide reputation for its scientific quality". Needless to say, the financial success depended very much on the unpaid cooperation of the Valentine family. The world-wide reputation was the result of Valentine's own breadth of interest and his editorial skill. Many a former student can testify to the illumination given by articles in the symposia for which the journal was noted and to the enthusiasm aroused by the more technical articles. It can be argued that the journal was Valentine's greatest contribution to psychology.

Those who wish to know what Valentine himself thought about the nature and methods of psychology can learn from the presidential address he gave to the British Psychological Society in 1948 under the title 'Some Present-day Trends, Dangers and Possibilities in the Field of Psychology'. It is naturally a popular treatment, but it illustrates his ability to compound wisdom with humour and common sense with insight. "The true psychologist", he said, "should indeed love his work for its own sake"; but "psychologists are also human beings (though that may surprise some people) and most of us are eager to see our work helping some great purpose".

Mr. Birch's 'List of Publications by C. W. Valentine' (*Brit. J. educ. Psychol.*, 1956, 26, Part 1) contains about seventy items, illustrating not only their wide range of subject matter but also versatility in technique. Valentine's most substantial book is *The Psychology of Early Childhood* (1942). It is a product of his systematic observation of his own five children and his extensive critical knowledge of writings about young children. It deals with their intellectual and emotional growth. The rapid development of new methods and psychological frameworks since it was written has not destroyed its value as a rich source of material.

Another important work is *Psychology and its Bearing on Education* (1950). This book made the author's eclectic approach to his subject and his lucidity as an expositor available to a much greater audience, designed as it was for students in training colleges and university departments of education. If it gives undue weight to experimental work on this side of the Atlantic it complements similar books published on the other.

Here one cannot deal in detail with the multifarious interests of Valentine's publications. In the early thirties he was concerned with the reliability of examinations and wrote a book with this title. The findings in the first half of it attracted much attention and set off other investigations, the results of which are now being seen in the changed attitude to the eleven plus examination. There are signs that the problems of university awards dealt with in the second half of the book will soon receive some attention. Clearly if huge sums are to be spent from the public purse on university education there will have to be conviction that selection and assessment are reliable. Valentine's inquiry was a pioneer effort, still to be followed up.

His interest in examinations was also expressed in *Examinations and the Examinee* (Birmingham Printers, 1938). One of his suggestions was the Quota Principle for the selection of children for grammar schools. This was put into practice by a former student, Mr. V. J. Moore, who was then Director of Education for the County Borough of Walsall. The essence of the scheme was to allocate a quota to particular primary schools and base the selection of the children for grammar schools on the school records and assessments. One could go on writing of Valentine's views: on the place of Latin in the school curriculum; of his restrained popularization of the psychology of the unconscious; of his early application of statistics to psychological experiments; of his intelligence tests; his views on discipline and bringing up children; on army morale and intuitive judgements. He treated all these topics and others shrewdly and lucidly, bringing to bear on them a scholarly and philosophical mind.

He was President of the Psychology Section of the British Association in 1930 and President of the British Psychological Society in 1947-48. In 1954 he was made an honorary member of the Société Française d'Esthétique. The University of Birmingham made him Professor Emeritus after he retired in 1946.

The humanity of his writings reflected his personality. He was invariably helpful to his students, infinitely removed from the professor whose response to a questioning student was, "I do not teach my subject, I profess it". He entertained students and staff at his house in Greenland Avenue, Selly Park, and his young colleagues greatly appreciated the opportunity of meeting their seniors at his table. He continued to lecture to the end of his life, enjoying this as well as watching cricket at Edgbaston. His friends and students will long remember the warmth of his personality and the encouragement they received from him.

W. J. SPARROW

To Professor Valentine this *Journal* owes a special debt of gratitude. Although he modestly disowned any claim to be himself a statistical psychologist, he was convinced from the very outset of his work that in educational psychology, as in most other branches, every conclusion should, so far as possible, be demonstrated or checked by rigorous statistical analysis. Since in the early days few readers of British journals were mathematically equipped, few British editors were willing to accept statistical papers. Valentine himself maintained that the more technical parts of the demonstration—the symbolic formulae, the algebraic proofs, the elaborate numerical tables—should be excluded from 'subject journals' and relegated to a 'journal of method', edited not for the general reader but for the research worker or academic teacher.

Accordingly, when a small group of British statistical psychologists proposed the publication of such a journal in this country, he strongly urged the Society to build up a fund for that purpose. The war intervened. But as soon as it was over the project was revived; and, when Valentine became President, he succeeded in getting the present *Journal* launched—largely on the basis of financial help from the authors interested and their various friends and benefactors. He assisted in drafting a statement of policy which gave the *Journal* a broad basis. His personal experience and advice were invaluable in drawing up plans for editing, printing, and publishing the new venture.

CYRIL BURT

BOOK REVIEWS

Personality Assessment: a Critical Survey. By P. E. VERNON. London: Methuen & Co., 1964. Pp. ix + 333. 42s.

Psychologists have had more than a passing interest in personality assessment since the 1930's, and this interest has greatly increased since the war, especially in the United States, despite growing disillusionment about specific assessment techniques. A comprehensive description and appraisal of this field has been sorely needed; this need has now been admirably met by Professor Vernon's newest book.

Staying close to the data contained in the articles cited in his 28-page bibliography—most of the important works in the field seem to be represented, with the surprising exception of Murray's ground-breaking *Explorations in Personality*—Vernon pragmatically examines and appraises each of the major approaches to personality assessment.

He begins by considering naïve observers' judgements of others, for their accuracy is a base-line against which the judgements of the presumably more sophisticated psychologist and psychiatrist can be compared. In addition, many of the factors that affect naïve observers' accuracy, the theories offered to explain their judging ability, and the ticklish methodological problems faced in quantifying judgements and evaluating their accuracy apply with equal force to clinicians' judgements.

Vernon then turns to assessments by clinicians. The clinical vs. statistical prediction controversy is brought up to date by his review of the reliability and validity of clinicians' judgements. He concludes, in agreement with Meehl, that predictions by clinicians are no more accurate than predictions by actuarial methods, or even by naïve observers. It is difficult, however, to accept Vernon's view that such a conclusion implies that interpretations based on depth psychology or psychoanalytic concepts are ineffective. His view assumes that the predictions of such variables as nosological categories, occupational success and psychotherapeutic change were solely, or at least mainly, based on depth-psychological interpretations, but it is not even known, for example, that most of the clinicians in these studies were psychoanalytically oriented.

After reviewing studies of projective techniques (mainly the Rorschach and TAT), he concludes that these techniques are not useful in personality diagnosis, and, in fact, "may detract from, rather than add to" ordinary case-history methods and objective tests, although they may have some value as exploratory instruments in psychotherapy or in measuring rather specific motives in research, e.g., McClelland's measurement of need achievement with the TAT.

The effectiveness of psychotherapy and counselling is also examined by Vernon, who sees it as another index of the validity of clinical assessments. The link between the two seems to be a tenuous one, at best, for the outcome of psychotherapy and counselling probably depends on many other things besides the clinician's accuracy in assessing his patient's personality—one can conceive of a therapist who possesses such accuracy but is incapable of modifying his patient's behaviour or producing insight in him.

In attempting to understand why clinicians' judgements seemingly go wrong, Vernon looks at the empirical evidence for some of the concepts of depth psychology presumed to underlie their judgements, and comes away advocating the alternative approach to personality offered by the self-theories of Rogers, Kelly and others. Vernon's choice of self-theory is pragmatic and is largely governed by his views that counselling guided by this theoretical orientation is relatively successful, while psychotherapy and assessments based on depth psychology are ineffective.

Personality appraisal with tests and questionnaires is not neglected. Vernon considers in detail the trait approach to personality measurement, exemplified by the factor analytic work of Eysenck and Cattell, and discusses at length such related problems as the test-taking attitudes (especially acquiescence and social desirability response styles) that are creating havoc with trait measurement and various conceptions of validity. The psychometric approach does not fare much better than the other assessment approaches: Vernon concludes that tests and questionnaires do not have enough validity to justify their use in individual diagnosis, and that their major weakness lies in their instability.

The author also presents a novel model of a personality system intended to integrate existing conceptions of personality. This system consists of the interactions between such characteristics of the person as his behaviour, dispositions, self-concepts and personal concepts of others; and such environmental characteristics as stimuli and the conceptual system of others.

Finally, Vernon discusses specific kinds of tests and other assessment techniques that he sees as useful in the light of the previous appraisals although not necessarily free of shortcomings. These techniques run the gamut from more or less traditional psychometric tools (e.g., self-report inventories, Q-sorts, situational tests, and the Semantic Differential) to novel approaches employing cognitive styles and somatotypes.

All but the most ardent psychometrician or clinician would have to agree, if only grudgingly, with most of Vernon's carefully reasoned and balanced conclusions. Still, it needs to be stressed that these conclusions refer to the techniques used at present in personality assessment, most of which were developed years ago and do not mirror our current sophistication and technology. A great deal is now known about why these techniques go wrong and how they can be improved, as Vernon observes, but, unfortunately, there have been few attempts to use this knowledge in improving them. It seems premature to discard projective techniques or other assessment tools before it is certain that such attempts to improve them would be fruitless. For example, the psychometrically sophisticated Holtzman inkblots, which were recently developed to eliminate many of the pitfalls that plague the original Rorschach, should demonstrate the inherent potential of this projective technique. Similarly, our present knowledge about the way that a personality inventory should be constructed and our technology (computers that can factor-analyse large numbers of items, and efficient analytic rotation procedures) would make it possible to construct, along the lines of the test construction programme described by Vernon, an inventory that is homogeneous, minimally affected by response styles, and free of other problems that interfere with the effectiveness of current inventories. Experience with such an inventory would be a better guide to the usefulness of personality scales than our accumulated experience with such anachronisms as the MMPI.

The book seems remarkably free of errors, but one worth noting was observed in the description of the Edwards Personal Preference Schedule (mistakenly identified as the Edwards Personality Preference Schedule). The text reports that the items in this forced-choice personality inventory were chosen, in part, on the basis of the items' correlations with the traits and with social desirability, but published reports of its construction indicate that no such correlations were used in selecting items. Instead, the items were chosen on two bases: that their content appeared to reflect the traits, judging simply from an inspection of the items, and that their rated social desirability was appropriate.

The manner of presentation could be improved. The prose is less than sprightly, there are some awkward and ambiguous sentences, some of the conclusions are so heavily qualified or otherwise obscured that they convey little meaning, and Vernon's own conception of validity and the term "notional criteria" crop up repeatedly but are never sufficiently defined or described. The book's organization is also distracting. There is considerable overlap in the contents of the Foreword and the first chapter, which both summarize the entire book, and the chapters on validity and the personality system are poorly located. The validity chapter appears near the end of the book, though the

perspectives that it offers on validity would have been useful in reading the preceding chapters that hinge on the question of validity. Similarly, the chapter on the personality system would be more appropriately located near the chapters that describe the depth psychology and self-theory approaches to personality.

The defects of this book detract only slightly from its considerable merits. The specialist in personality assessment will find it to be a useful reference, and a source of interesting and stimulating insights; other psychologists and students of psychology will find it to be a unique introduction to an important field of psychology.

LAWRENCE J. STRICKER

Famous Problems and Other Monographs. Edited by R. C. ARCHIBALD. New York: Chelsea Publishing Co., 1962. Pp. 321.

This volume, intended primarily for students, consists of four mathematical monographs, and forms the latest publication in the useful series issued under the title of 'Chelsea Scientific Books'. To statistical psychologists the contribution of greatest interest is the work published forty years ago by W. F. Sheppard and entitled 'From Determinant to Tensor'. It is in effect an introduction to matrix algebra, with an eye to statistical problems, like those encountered by the psychologist. Unfortunately its relevance is somewhat concealed by the peculiar terminology and notation which Sheppard insisted on using.

Among psychologists Sheppard seems to be known solely from the correction formulae which bear his name and the valuable tables of the normal distribution which appear (more or less abridged) in almost every statistical textbook. He was, however, one of His Majesty's Inspectors of Schools, and in that capacity played an important part in the theoretical study and practical applications of mental tests. I personally am glad of an opportunity to record the deep debt of gratitude which I owe to his friendship and inspiration.

Before Sheppard's monograph was published all attempts at proving the algebraic formulae required for factor analysis were carried out by Pearson's method of writing out determinants and other arrays in full. The change to matrix algebra, first suggested by Sheppard, enabled a proof which formerly occupied two or three pages to be condensed to little more than two or three lines—with a great increase in lucidity and a great reduction in the expense of printing. In adapting Sheppard's suggestion, I ventured to substitute Cullis's standard notation; Hotelling later on, in his proof of Pearson's method of 'principal components', preferred to employ tensor-notation, similar to that here used and explained by Sheppard.

In the present reissue Sheppard's monograph is followed by one from Major Macmahon, entitled 'An Introduction to Combinatory Analysis'. The clear and elegant treatment will make it useful for those interested in problems of probability. It should, however, be supplemented by a study of the more recent work on *Combinatorial Chance* by Professor David and Dr. Barton.

The publication ends with three lectures on Fermat's Last Theorem, given at Birkbeck College by Dr. Mordell. In or about 1637 Fermat scribbled in the margin of a book the statement that the equation $x^n + y^n = z^n$ ($n > 2$) cannot be satisfied by any integer, adding that he had "discovered a truly remarkable proof which this margin is too small to contain". What was his 'proof'? As has frequently been noted, when Fermat had derived a formula by induction only, he always conscientiously stated that he could *not* prove it: (e.g., the famous formula for primes, $2^{2^n} + 1$, which Euler later showed breaks down for $n=5$ since 641 is then a factor). In discussions of the 'Last Theorem' it is now customary to distinguish two cases—a point which Dr. Mordell does not bring out. For 'Case I' (which was

probably the one Fermat had in mind) it is generally agreed that the theorem is in fact true. Yet it is surely a point of interest to the psychologist that almost every mathematical genius in every civilized country has attacked the problem, and yet failed to rediscover Fermat's "truly remarkable" proof, or to win the prize of 100,000 marks offered by the Academy at Paris for a successful solution. 'Case II' turns out to be still more elusive.

For those who enjoy mathematical puzzles the opening treatise by Felix Klein on 'Three Famous Problems' is the most fascinating of all. Klein succeeds in explaining each 'problem' in an exceedingly simple style. To follow his discussion even an acquaintance with the calculus is unnecessary. The three problems are (i) how to duplicate the cube (the ancient Delian problem), (ii) how to trisect an arbitrary angle, and (iii) how to square the circle. The last involves the geometrical construction of π . Archimedes showed that the ratio of the circumference of a circle to its diameter lies between $3\frac{1}{7}$ and $3\frac{10}{71}$. Tsu Chung-Ching (5th century) gave the approximation $3\frac{16}{113}$ which is correct to six decimal places: since the fraction = $4^2/(7^2+8^2)$ this excellent approximation can be effected by Euclidean methods. However, with straight edge and compasses alone (as has long been recognized), an *exact* construction is impossible. Nevertheless, as Klein points out, a simple but ingenious instrument, called an 'integrator', has been devised by a Russian engineer which will do the trick.

The editor's supplementary notes, and the histories of the various issues discussed, add still further to the interest of the volume.

CYRIL BURT

Computers and Thought. Edited by E. A. FEIGENBAUM and J. FELDMAN. New York: McGraw Hill, 1963. 63s.

We have all speculated at some time about whether or not machines can be said to be intelligent. In the popular view the outward similarity of the behaviour of computers to that of humans is taken as evidence of 'machine intelligence'. More cautious people have been inclined to scoff at the idea, pointing out that machines can do only 'what we tell them to'. In most applications of computers this is clearly true. A program is prepared for the solution of a well-defined problem, with a specific set of instructions for every possible eventuality. The human programmer has, or hopes he has, thought of every possible path that the program may take. When he loads the program into the computer, we may say that he has made a special-purpose machine for the solution of a particular problem.

On the other hand, there are problems where the number of possible paths is very large, and the selection of any one path is heavily dependent upon the current state of the machine. It is impossible for the human programmer to consider every possible state in such problems. Despite this difficulty some progress has been made in the use of computers in this kind of situation: in the recognition of handwritten characters, the proving of geometrical theorems and playing chess and draughts. In these cases the machine can be thought of as performing processes which the programmer did not build into it. So the possibility exists for a machine to exhibit behaviour which could be described as intelligent according to some objective test.

The papers reprinted in the book *Computers and Thought* represent an important part of a growing body of work on these problems of machine intelligence by scientists from several disciplines. The book is divided into two parts. The first part consists of papers which make no claim to simulate any actual brain mechanisms, while the second part contains several papers which throw a new light on such subjects as rote learning and binary choice behaviour.

The collection begins with an important paper by A. M. Turing. He proposes a procedure for assessing machine intelligence, in which an interrogator must try to distinguish between a human and a machine by the answers to questions transmitted by teleprinter. This is followed by a very lucid exposition of the mechanistic argument for intelligent

behaviour in machines. It is noteworthy that Turing, to whom we owe the theoretical foundations of computation, took such a firm point of view as early as 1950.

Apart from Turing's paper and the two survey articles with which the book concludes, all the papers describe actual machine experiments, in many cases accompanied by protocols of machine behaviour. Although none of them come near to satisfying Turing's test, the results from those within a sufficiently restricted problem area are impressive for the newcomer to this field. The remark of the draughts champion defeated by A. Samuel's learning program is illuminating: "In the matter of the end game, I have not had such opposition since 1954, when I lost my last game."

The problem of increasing the generality of problem-solving mechanisms is attacked by Newell, Shaw and Simon in the General Problem Solver, which is an algorithm for generating and seeking sub-goals in a general problem environment. But there are many aspects of human problem-solving behaviour besides goal-seeking which require greater formalization before they can be included in a mechanical model. Not least is the ability to form concepts and this may well be the crucial problem in the study of machine intelligence.

The editors, who are psychologists, have produced a collection of papers admirably balanced between the empirical approach to machine intelligence and the contribution of the study of computers to the theoretical understanding of human behaviour. The introductory passages to the various sections of the book are first-class discussions of their subjects and provide a thread of continuity throughout. The non-programmer may find the technical details of some of the papers onerous, but the method is clear in most cases without these details. The contributions were originally published in a widespread section of journals and proceedings, and in many cases the authors have brought their papers up to date for this book. There is an exhaustive bibliography which will be invaluable to workers in the field. Anyone who is at all curious about machine intelligence should read this book.

GEORGE F. COULOURIS

Handbook of Mathematical Psychology, Vol. I. Edited by R. D. LUCE R. R. BUSH and E. GALANTER. New York and London: Wiley, 1963. Pp. xii+491. 80s.
Readings in Mathematical Psychology, Vol. I. R. D. LUCE, R. R. BUSH and E. GALANTER. New York and London: Wiley, 1963. Pp. 535. 68s.

E. G. Boring once remarked that "the work of the mathematicophiles may perhaps be explained by the fact that the elaborate and arduous mathematical method distracts attention from equal care in the selection and experimental attainment of data". Today it is still easy to find psychologists who regard this comment as apt. On the other hand, it must be admitted that the 'mathematicophiles' now make up a larger proportion of psychologists than hitherto and they cannot be dismissed quite so summarily. Indeed, there has been an almost explosive growth in the number of psychological papers relying on mathematically formulated arguments. The mathematicophiles are multiplying.

But to what effect? Is it now necessary to provide a good background in mathematics for psychology students because otherwise they will be prevented from understanding important research and contributing to it? The present volume should help to provide answers to questions of this kind. It is the first in a series consisting of three handbooks of mathematical psychology and two accompanying volumes of selected readings, the latter taking the form of anthologies of reprints.

The editors have been giving courses on mathematical approaches to psychological problems at the University of Pennsylvania for several years. Their original intention was to put the relevant research material into a single volume which would be suitable as a textbook for undergraduate and graduate courses. However, the contributions of their authors were such as to lead to the undertaking of the present more ambitious enterprise. The first volume, reviewed here, is concerned mainly with measurement and psychophysics.

The second, which has also appeared, is devoted to mathematically formulated theories of learning and social behaviour. The last, we are told, is a more miscellaneous work and deals with the senses, learning, preference and mathematics *per se*. The first volume of the ancillary readings covers measurement, psychophysics and learning, and is an interesting and stimulating collection of reprints. The remaining topics will be covered in the second volume. There remain several areas of psychology where mathematics has flourished which are not discussed in the present series. In three instances this was intentional: information theory, scaling and factor analysis were considered to have been fully described elsewhere. On the other hand, the theory of mental tests was not covered because the editors did not manage to find a willing author. One cannot but regret this omission. It is an area in which there are many technical complexities and an authoritative essay on approaches now current would have been extremely welcome.

As far as the first volume is concerned, the mathematical demands on the reader are not very great. Given an intention to understand the arguments, a point somewhere on the road from 'O' to 'A' level school mathematics should be an adequate mathematical preparation. However, several of the chapters, particularly Six and Eight, do presuppose a considerable knowledge of probability theory and mathematical statistics. But the greatest obstacle to understanding for the average 'mathematicophile' is likely to be an absence of sufficient knowledge of logic to cope with the first chapter and a related unfamiliarity with the kinds of detailed axiomatic statements of mathematical notions which occur from time to time throughout the book. The latter, however, do not present an insurmountable problem and the reader is often given considerable guidance by the authors. However, although the first chapter may be read, so to speak, between the axioms, a proper understanding of it does demand some facility in formal logic. In the reviewer's opinion the Handbook is best approached by means of a small seminar in which various mathematical and statistical intuitions are represented: a logician may be useful.

The above comments are not meant to be adverse to the substance of the opening chapter by Suppes and Zinnes on 'Measurement Theory': the ideas it contains are expressed in a clear and unambiguous way. The authors approach their topic by formulating two fundamental problems: (1) the justification of the assignment of numbers to objects or phenomena; (2) the specification of the extent to which this assignment is unique. The first problem is dealt with by precisely characterizing the formal properties of empirical operations used in a measurement procedure and showing that they are isomorphic (perhaps homomorphic) to the operations and relation in an appropriately selected system of numbers.

'Isomorphic' here refers to a one-to-one correspondence between, on the one hand, a total empirical system of measured objects and the relations among these measurements and, on the other hand, the elements and relations which characterize the selected descriptive number system. The less stringent homomorphic correspondence between the empirical and numerical systems is mentioned because there arise cases where two objects have the same measurement value assigned to them. The approach may seem so broad as to make 'measurement' synonymous with 'theory', but the authors stress the point that the measurement procedure is only likely to be of value when the numerical system employed is formed of simple and familiar relations. They attach no particular value to the operation of addition which some writers have considered essential to the notion of measurement. Neither do they consider that the carrying out of certain empirical operations on the measured objects is a criterion which must be satisfied before measurement can be said to have been achieved. The whole emphasis is placed upon the correspondence of the relations in the empirical system of objects to those in the system of numbers representing the scale of measurement.

The question of the uniqueness of the numerical system used in any situation is answered by considering the different numerical systems which could be successfully used to describe the same system of objects. The particular kind of measurement scale is then determined by the nature of the transformations which map any of the alternative scales on

to one another. Thus for so-called 'interval scales', if x measures an object on one scale and y measures it on another then it will be found that $y = a + bx$ for all objects. Hence interval scales are said to be 'unique up to a linear transformation', since the assignment of numbers can be manipulated in this way without disturbing the faithful representation of empirical relations. A variety of measurement systems are examined including the semiorders of Luce (where it is possible that $a = b$ and $b = c$ but $a > c$: as in the discrimination of very similar stimuli). The authors then go on to consider derived measurement. An example from the physical sciences is density; from psychology, the representation of paired comparisons data by the Bradley-Terry Model. In the latter the relative frequency with which object i is preferred to object j , P_{ij} , is expressed in the form $P_{ij} = \pi_i / (\pi_i + \pi_j)$. The π 's are derived measures of preference for the various objects.

Although the formal analysis of measurement is both thorough and intellectually stimulating, the chapter lacks, apart from one or two illuminating asides, any adequate discussion of the measurement problem in relation to actual psychological observations. In a handbook these practical aspects should have been covered.

The next four chapters essentially constitute a monograph within the Handbook. Chapter Two is an essay on the characterization of experiments on choice which serves mainly to introduce the notation adopted in the following three. Some useful distinctions are made: particularly that separating the rules that an experimenter uses to allot rewards and punishments to responses (outcome function) from those governing the specification of the response which is, in a manner of speaking, the correct one to make in a given situation (identification function). For example, in a study in which the subject is trying to detect very weak signals, if he fails to detect it the identification function specifies that this circumstance requires the same response as would usually be given in the absence of stimulation. On the other hand, the experimenter may be heavily rewarding all detections and this outcome may determine the subject to report the presence of a stimulus since this may maximize his expected gain or optimize his performance in some other way. Like all classification schemes this one reveals some kind of experiments which have not yet been conducted and someone might be tempted to investigate them.

The third chapter, on 'Detection and Recognition' by Luce, is one of the best in the Handbook. It is mainly about situations in which a subject is presented with one of several kinds of stimulation and has to decide which it was. Greatest attention is given to cases where there are two kinds of stimulation: that from the background 'noise' of a threshold study and that arising when a weak signal is added to this background. Luce points out that a satisfactory analysis of this kind of experimental task cannot regard the subject's performance as a direct function of his sensitivity alone. It is also necessary to indicate how his performance will be affected by motivation; for example, by the rewards and punishments contingent upon correct detections and false alarms. Of the three hypotheses that are considered, the first is the signal detectability model in which the subject is represented as calculating the likelihood of the present stimulation according to the alternative hypotheses of the source being 'noise' and 'noise plus a signal', taking the ratio of these, and announcing the presence of a signal when this ratio exceeds some criterial value. Here, sensitivity is measured by the estimable distance between the means of the likelihood ratio distributions resulting from the two alternatives. Performance varies by an adjustment to the position of the criterion on the likelihood ratio axis. Luce then considers an extension of his treatment of detection in terms of the choice model which was the subject of his monograph 'Individual Choice Behaviour'. In this scheme, sensitivity is measured by a response-defined measure of the similarity between the two sources of stimulation, and performance may be modified by a bias in favour of making one or other of the responses. Finally, a threshold model of detection is presented. However, Luce has extended the analysis of this traditional model by supposing that there is a definite probability that the threshold will be exceeded when no signal is present. The subject may still adjust his performance: by falsely reporting that he detected a signal on trials

on which the threshold was not exceeded, or falsely denying detection when it was exceeded. One result of this generalization is that the theory can now be used to describe satisfactorily certain gross features of the function relating the probability of detecting signals to the probability of falsely designating noise as a signal (the ROC curve).

Luce shows how the schemes fare in a variety of situations and altogether the chapter is a very good guide through the problems of this area of psychophysics. On the experimental side, he is perhaps not sufficiently critical of his own theory. In particular, little is said about the use of a rating technique in detection experiments which suggests that subjects can order sensory input even at the low intensity levels which would correspond to the threshold of traditional theories. This is a finding which would seem to be hard to represent within the structure of a theory which is stated in terms of responses biases.

Chapter Four is on 'Discrimination' and was written by Luce and Galanter. It is concerned mainly with discriminability scales: Fechner's original psychophysical law; Thurstone's 'law of comparative judgement'; and the Bradley-Terry-Luce scale for representing paired comparison data. In all cases these are derived measurements based on the notion that confusion between stimuli may be used to scale the psychological distance between them. The emphasis on scaling in this chapter will make it of less interest to most psychologists than the one on detection. The authors claim that there are relatively few data that are suitable to test existing mathematical theories in this area of psychophysics. It may well be that the authors are considering the wrong kinds of theories of discrimination. If we still do not have any useful account of different ways in which stimulus patterns can be discriminated, then a useful chapter would have been one giving a careful mathematical analysis of the most common psychophysical techniques used in the study of discrimination.

The same authors make a general survey of most other kinds of psychophysical scales in Chapter Five. This is much more interesting than the treatment of discriminability scaling because here they attempt to consider the kinds of operations which the different scaling techniques would seem to demand from a subject. In particular there is a very careful analysis of some of the issues raised by the magnitude estimation method championed by Stevens. There are clear indications of an interaction between the original views of the authors and Steven's critical comments on an earlier draft of the chapter. A very spirited attempt is made to determine the extent to which any true psycho-physical function relating stimulus to psychological magnitude can be separated from responses biases and motivational variables operating in the typical laboratory experiment. It is certain that in the debate which will undoubtedly continue about the 'power law' the pages of this chapter will be often thumbed.

Chapter Six on 'Stochastic Latency Mechanism' by McGill is regarded by the reviewer as something of a *tour de force*. It is a systematic examination of various systems which could be postulated to account for the delay between the presentation of a stimulus and the ensuing response. General techniques for obtaining the characteristics of such systems are first clearly outlined. Then starting from the simplest model of waiting times—the Poisson process which leads to an exponential distribution of response latencies—successive complications are studied; adding further independent processes to the chain of events filling the stimulus-response interval; making the response the first to occur in several similar parallel systems; and so on. Although the chapter would have been even better if it had provided a systematic analysis of temporal aspects of choice, McGill does indicate how this could be contrived. Indeed, there are a host of theoretical ideas in the modest and lucid commentary which accompanies the mathematical formulae. Illustrations of the various schemes are nicely selected from a wide range of experimental results and there can be few psychologists using response time as a datum who will not find this a very illuminating chapter. It is to be hoped that this essay may be the sketch plan for a book.

Newell and Simon contribute the next chapter on 'Computers in Psychology'. Unlike the others which are introductions to the actual mathematical techniques used in

particular topics, this one is confined, almost unavoidably, to a survey of what has and might be done in the way of computer simulations of behaviour. Although written by well-known computer enthusiasts it gives a very sober appraisal of the problems of comparing computer simulations with real behaviour. In the context of the Handbook, with its over-all emphasis on analytic methods and explicit mathematical formulae, the most challenging aspects of this chapter for psychologists will be those concerned with so-called 'information processing programs'. In these, certain processes are specified for dealing with symbols of all kinds, not just numerical ones. Such schemes can be used to simulate quite complex cognitive processes, but the cost of this possibility of a rigorous theoretical approach to complex psychological phenomena is that it is not usually possible to make general inferences about the performance of such systems and it becomes difficult to test them empirically. The clear discussion of problems of this kind and the variety of simulations presented in this chapter make it an excellent aperitif to the computer literature.

The last chapter by Bush is a short one, 'Estimation and Evaluation'. The problems of the first process are well understood and a simple illustrated account of the various available procedures is provided. The evaluation of models is only lightly touched upon and is perhaps too closely linked with an accurate representation of the experimental data from any given situation. Even when writing of the virtues of an hypothesis which covers several different situations, Bush places emphasis upon parameter invariance. There must, surely, often be grosser features of a model which argue for or against its acceptance and make it unnecessary to go into the details of testing for the stability of parameters across various situations.

In conclusion it can be said this volume clearly demonstrates that mathematics can be usefully employed in the analysis of a wide range of psychological problems and at the same time it provides a useful illustrated guide to some areas where this analysis has been most thoroughly conducted. Excepting the chapter on computer simulation, that by McGill on latency mechanisms and part of Luce's chapter on detection, it could nevertheless be argued that too much emphasis is placed upon a rather formal description of experimental situations. The approach to a problem is often made from a mathematician's point of view and we are not always given a mathematical analysis of plausible psychological hypotheses. The title of Handbook is not entirely appropriate; it is probably too early for any work to achieve that status, and this one retains the flavour of the introductory textbook which the editors originally had in mind. But an introductory textbook of mathematical psychology was sorely needed and the editors and authors are to be congratulated on their achievement. The reviewer expects that many, like himself, will want to use this work as the basis for postgraduate seminars for some time to come.

R. J. AUDLEY

The Logic of Preference. By G. H. VON WRIGHT. Edinburgh: The University Press, 1963. Pp. 68. 10s. 6d.

This short study is an expanded version of lectures given in the University of Edinburgh in 1962. The author is not concerned with utility or "measurable value" nor with probability. He acknowledges Ramsay's influential essay 'Truth and Probability' (1931) and Halldén's 'On the Logic of Better' (1957), but takes a slightly different standpoint. His object is to study preference from a formal logical point of view, on the grounds that preference can, as logically prior, be studied independently of utility and probability. He comments that "authors on these topics usually take for granted certain logical features of preferences—such as asymmetry and transitivity—and then hasten on to utility (and/or probability) . . . the temptation is worth resisting for the sake of first creating a systematic formal theory of preference".

Readers are assumed to be familiar with elementary propositional calculus. Using P as a relational operator, and p, q, r as primitives, von Wright—after limiting himself to preference in a single instance, and making the distinction generally accepted between choice as observable behaviour and preference as an underlying determinant of behaviour—obtains a number of conclusions pertinent to preference not involving risk. In his system pPq is “tantamount” to saying $(p \cdot \sim q)P(\sim p \cdot q)$ where the antecedent and consequent of this latter relation are taken to be contemplated possible changes in the subject's present condition. By considering preferences among the triples p, q and r he obtains a distinction between conditional and unconditional preferences; this raises difficulties at the intuitive psychological level which are circumvented by postulating a universe of discourse which is the set of all variables $p, q, r \dots$ and then ‘embedding’ preferences in this universe, to reflect what von Wright calls the ‘*ceteris paribus*’ quality of preferences. Examining permissible transformations of P -expressions, von Wright identifies P -tautologies and shows that what is an equivalence in propositional logic may not be so in P -logic. He states “expressions which are probably equivalent in the logic of propositions are not, without restriction, intersubstitutable in expressions of the logic of Preference. The restriction is that the substitution must not introduce new variables into the atomic P -expression in which the substitution takes place.”

The other major point of interest to psychologists is that von Wright is obliged to distinguish between I (indifference) and E (equivalence), so that value-equality is stronger than indifference; states which are value-equal are necessarily indifferent; but states which are indifferent between themselves are not necessarily value-equal. This is shown to follow from his definitions of I and E .

Miller (*Mathematics and Psychology*, 1964) has recently pointed out that a shift historically in psychologists' attitudes to normative theory has taken place; such theories are now acceptable as reference models against which to compare actual behaviour, even when they are known to be in some degree false. Formerly they tended to be out of favour, which in part accounts for the difference in development of economic and psychological theories of choice behaviour. Von Wright does not help us to see how his work might be used as a normative theory of behaviour, in part because of his peculiar use of words (e.g. ‘anthropological’ when he means ‘psychological’) and partly because, although he feels his theory is fundamental to economic, aesthetic and moral choice, it is too simple to describe economic choice. We have too little data to assess its relevance to moral choice, but a comparison with Piaget's *Logic and Psychology* (1953) is of interest for those concerned with the genetic development of ethical systems. The reviewer feels it may be directly relevant to the comparative analysis of aesthetic choice behaviour with and without irrelevant alternatives present, but again there is evidence that aesthetic preferences are not readily described by the proposed system.

Von Wright's difficulties with indifference and transitivity are closely similar to those faced by previous writers such as Luce or Arrow, who have constructed stochastic choice models from preference-utility-probability axioms, and much remains to be done if his book is to become the basis of experimental research. The book is less satisfying than his earlier analysis of modal logic, but it may be read with profit by psychologists who find von Neumann and Morgenstern too much to swallow.

R. A. M. GREGSON

INDEX TO VOLUME XVII

ARTICLES

<i>Author</i>	<i>Title</i>	<i>Page</i>
Adcock, C. J.	Higher-Order Factors	153
Burt, C.	Consciousness and Space Perception	77
Burt, C.	Is Intelligence Normally Distributed? A Reply	89
Burt, C.	The Stability of Factors	177
Coan, R. W.	Theoretical Concepts in Psychology	161
Cowan, J. D. and Arbib, M. A.	Obituary: Norbert Wiener	90
Das, Rhea S.	Item Analysis by Probit and Fractile Graphical Methods	51
Emmett, W. G.	Secondary School Selection	86
Gregson, R. A. M.	Fitting a Linear Trace Decay Model to Taste Comparisons and Preferences	137
Hendrickson, A. E. and White, P. O.	Promax: a Quick Method for Rotation to Oblique Simple Structure	65
Keats, J. A.	A Method of Treating Individual Differences in Multi-Dimensional Scaling	37
Kristof, W.	Testing Differences between Reliability Coefficients	105
Lawley, D. N. and Maxwell, A. E.	Factor Transformation Methods	97
Luce, R. D.	Asymptotic Learning in Psychophysical Theories	1
Peaker, G. F.	Is Intelligence Normally Distributed?	89
Shallice, T.	The Detection of Change and the Perceptual Moment Hypothesis	113
Sparrow, W. J.	Obituary: Charles Wilfrid Valentine	180
Taylor, J. G.	What is Consciousness?	71
Treisman, M.	The Effect of One Stimulus on the Threshold for Another	15
Vernon, P. E.	A reply to Mr. Emmett's Letter	88

BOOK REVIEWS

<i>Author</i>	<i>Title</i>	<i>Page</i>
Archibald, R. C. (Ed.)	Famous Problems and Other Monographs	185
Buckland, W. R. and Fox, R.	Bibliography of Basic Texts and Monographs on Statistical Methods	94
Cohen, J. (Ed.)	Readings in Psychology	96
Feigenbaum, E. A. and Feldman, J. (Eds.)	Computers and Thought	186
Halmos, P. and Iliffe, A. (Eds.)	Readings in General Psychology	96
Humphrey, G. (Ed.)	Psychology through Experiment	93
Jöreskog, K. G.	Statistical Estimation in Factor Analysis	95
Luce, R. D., Bush, R. R. and Galanter, E. (Eds.)	Handbook of Mathematical Psychology, Volume I	187
Luce, R. D., Bush, R. R. and Galanter, E. (Eds.)	Readings in Mathematical Psychology, Volume I	187
Ramakrishna, B. S. et al.	Some Aspects of the Relative Efficiencies of Indian Languages	93
Vernon, P. E.	Personality Assessment: a Critical Survey	183
Von Wright, G. H.	The Logic of Preference	191



INDEX TO VOLUMES I TO XIV

An index to the first fourteen volumes of the *British Journal of Statistical Psychology*, classified according to authors and subjects, is now available. As proprietors of the journal the British Psychological Society wish to take this opportunity of expressing their gratitude to Mrs. Beatrice Warde to whose generosity and expert assistance the compilation and printing of the index is due.

OCCUPATIONAL PSYCHOLOGY

Editor: ALEC RODGER

JULY-OCTOBER 1964

VOLUME 38, Nos. 3 AND 4

CONTENTS

- Experiments on the Use of Programmed Instruction to Increase the Productivity of Training.
By D. WALLIS
- The Influence of Assembly Line Organization on Output, Quality and Morale.
By H. G. VAN BEEK
- Some Characteristics of Respondents, Partial Respondents and Non-Respondents to
Questionnaires on Job-Satisfaction.
By MARY SPEAK
- Studying the Employment and Training of a National Sample of 17-year-olds.
By DAVID M. NELSON
- Sorting Sheep from Goats with One Yardstick.
By R. A. M. GREGSON
- Preferred Patterns of Duty in a Flexible Shift-Working Situation.
By GWYNNE THOMAS and SYLVIA SHIMMIN
- Research into Problems of Peace and War.
By E. ANSTEY
- Book Reviews.

Annual Subscription 40 shillings, overseas 50 shillings

National Institute of Industrial Psychology, 14 Welbeck Street, London, W.1

BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

Vol. XVII, Part 2

November 1964

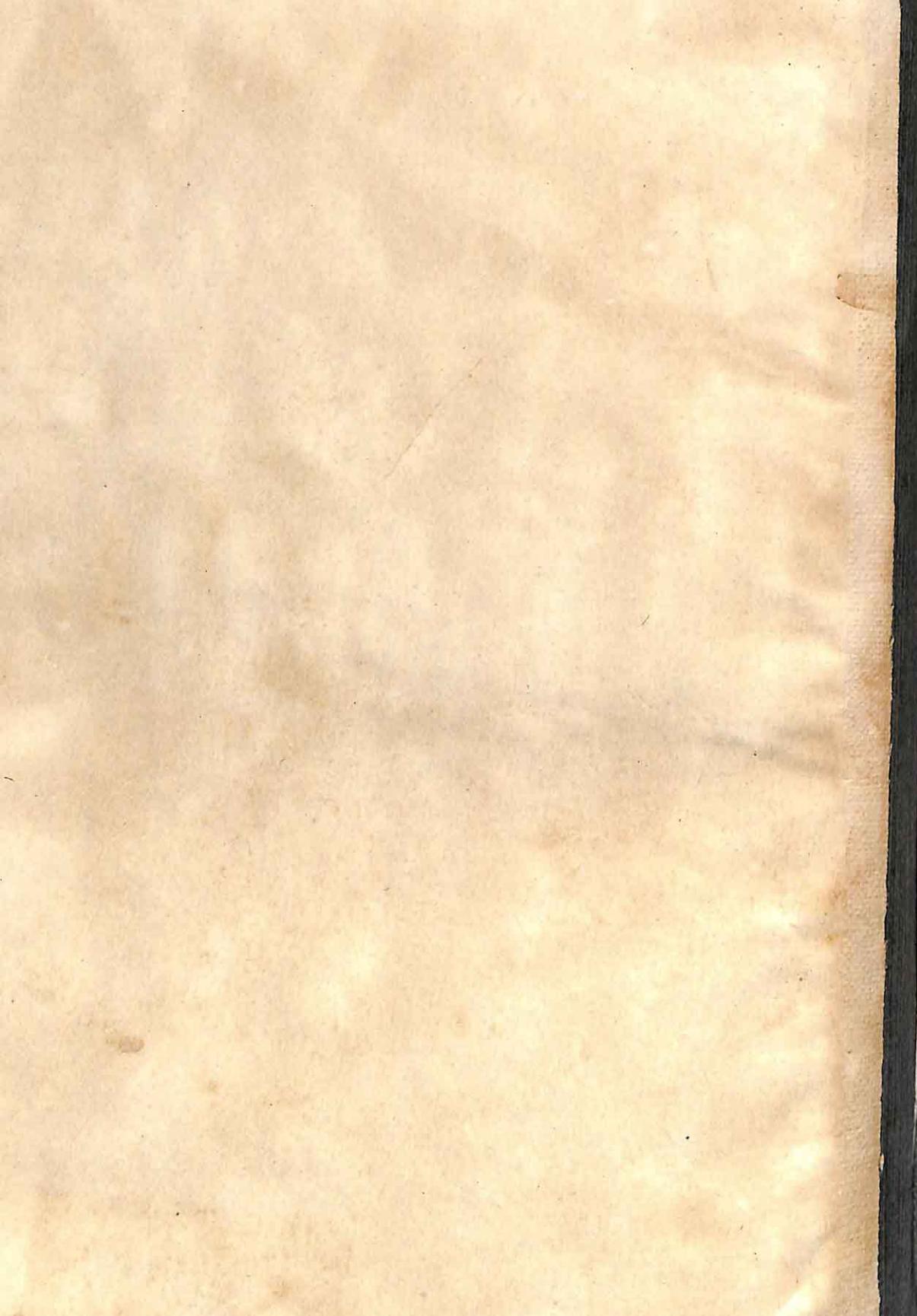
CONTENTS

	<i>Page</i>
LAWLEY, D. N. AND MAXWELL, A. E.	Factor Transformation Methods 97
KRISTOF, WALTER	Testing Differences between Reliability Coefficients 105
SHALLICE, T.	The Detection of Change and the Perceptual Moment Hypothesis 113
GREGSON, R. A. M.	Fitting a Linear Trace Decay Model to Taste Comparisons and Preferences 137
ADCOCK, C. J.	Higher-Order Factors 153
COAN, RICHARD W.	Theoretical Concepts in Psychology 161
NOTES AND CORRESPONDENCE	
	The Stability of Factors 177
OBITUARY	Charles Wilfrid Valentine 180
BOOK REVIEWS 183









MADE BY THE DISABLED
REHABILITATION - INDIA

